

Masterarbeit

Kim Radmacher

DSP basierende Bitratenreduktion von
Audiosignalen auf der Grundlage von MPEG-1

Kim Radmacher
DSP basierende Bitratenreduktion von
Audiosignalen auf der Grundlage von MPEG-1

Masterarbeit eingereicht im Rahmen der Masterprüfung
im gemeinsamen Masterstudiengang Mikroelektronische Systeme
am Fachbereich Technik
der Fachhochschule Westküste
und
am Department Informations- und Elektrotechnik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr.-Ing. Hans Peter Kölzer
Zweitgutachter: Prof. Dr.-Ing. Hans-Dieter Schütte

Abgegeben am 29.06.2015

Kim Radmacher

Thema der Masterarbeit

DSP basierende Bitratenreduktion von Audiosignalen auf der Grundlage von MPEG-1

Stichworte

MPEG, Bitratenreduktion, Audiokompression, Psychoakustik, Filterbank, FFT, Maskierung, Codierung, DSP, Frequenzgruppen, SPI

Kurzzusammenfassung

Diese Arbeit befasst sich mit der Reduzierung der Bitrate von Audiosignalen. Als Grundlage dient das MPEG-1 Audio Layer I Verfahren. Zunächst wird die nötige Theorie betrachtet und diskutiert. Im Anschluss folgt die Entwicklung und Simulation eines Encoders und eines Decoders. Das simulierte Konzept dient als Vorlage für eine Umsetzung auf dem DSP TMS320C6713B. Anhand abschließender Messungen wird das System beurteilt. Dabei kommen auch Hörtests zur subjektiven Beurteilung zum Einsatz.

Kim Radmacher

Title of the paper

DSP based bit rate reduction of audio signals based on MPEG-1

Keywords

MPEG, bit rate reduction, audio compression, psychoacoustics, filterbank, FFT, masking, coding, DSP, critical-band rate, SPI

Abstract

This thesis studies the reduction of the bit rate of audio signals, based on the MPEG-1 Audio Layer I process. First the required theories will be considered and discussed. Then both an encoder and decoder will be developed and simulated. The simulated concept subsequently serves as a template for an implementation on the DSP TMS320C6713B. Finally, the system will be rated based on readings, including aural tests for subjective evaluation.

Inhaltsverzeichnis

Tabellenverzeichnis	6
Abbildungsverzeichnis	7
Abkürzungsverzeichnis	11
1. Einleitung	12
1.1. Motivation	12
1.2. Stand der Technik	13
1.3. Aufgabenstellung	14
1.4. Gliederung	15
1.5. Notation	15
1.6. Geräteliste	16
2. Psychoakustik	17
2.1. Absolute Hörschwelle	17
2.2. Maskierung und Mithörschwelle	19
2.2.1. Zeitbereichsmaskierung	19
2.2.2. Frequenzbereichsmaskierung	20
2.3. Frequenzgruppen	22
3. Filterbank	26
3.1. Zweikanal Filterbank	26
3.1.1. Analyse-Filterbank	27
3.1.2. Synthese-Filterbank	28
3.2. Verzerrungs- und Aliasfunktion	30
3.3. Perfekte Rekonstruktion	30
3.4. M-Kanal Filterbänke	31
3.4.1. Baumstruktur	31
3.4.2. Parallelstruktur	32
3.5. DFT-Filterbank	33
3.6. Cosinus-modulierte-Filterbank	37
3.6.1. Optimierung	46

4. MPEG-1 Audio Codierung	49
4.1. Filterbank	51
4.2. Skalenfaktoren	56
4.3. FFT	57
4.4. Psychoakustisches Modell	59
4.5. Dynamische Bitzuweisung	70
4.6. Quantisierung	76
4.7. Rahmenbildung	77
4.8. Decodierung	78
4.9. Simulationsergebnisse	79
4.9.1. Spektrogramme	79
5. Implementierung	84
5.1. Verwendete Hardware	84
5.2. Datenübertragung	86
5.3. Filterbank	91
5.4. Audio Encoder	95
5.5. Audio Decoder	101
6. Messungen und Tests	102
6.1. Echtzeit	102
6.2. Amplitudengänge	102
6.3. Spektrogramme	104
6.4. Hörtest	109
7. Zusammenfassung und Ausblick	111
A. Mathematische Zusammenhänge	113
A.1. Polyphasenrepräsentation	113
A.2. Dezimation und Interpolation	116
B. Spektrogramme	119
B.1. Simulation	119
B.2. DSP Messungen	123
C. Inhalt des Datenträgers	127
Literaturverzeichnis	128

Tabellenverzeichnis

1.1. Notation	15
1.2. Verwendete Geräte	16
2.1. 24 Frequenzgruppen nach Zwicker	23
4.1. Skalenfaktoren	57
4.3. Indices k und korrespondierende Frequenzen sowie Werte der absoluten Hörschwelle LT_q und Tonheit z in Bark	62
4.4. Bitzuweisung	72
4.5. Zugewiesene Bits je Teilband (Simulation)	76
4.6. Verwendete Musikstücke für Messungen und Hörtests. Neben dem Namen der Stücke sind Interpret (ggf. Dirigent), Musikrichtung und die zeitliche Position eingetragen. Die Musikstücke sind durch die Notation $\mathcal{S}^{Bezeichnung}$ gekennzeichnet.	79
5.1. Zugewiesene Bits für die 32 Teilbänder mit Differenz zur Simulation	99
6.1. Messungen der totalen Berechnungsdauer für Echtzeitbedingung	102
6.2. Ergebnisse des Hörtests	110

Abbildungsverzeichnis

1.1.	Prinzipschaltbild des Gesamtsystems	14
2.1.	Absolute Hörschwelle nach E. Terhardt	18
2.2.	Verdeckungs- bzw. Maskierungseigenschaft des menschlichen Gehörs. Nicht-simultane Vormaskierung (<i>engl. pre-masking</i>), Nicht-simultane Nachmaskierung (<i>engl. post-masking</i>) und simultane Maskierung. Aufgetragen ist der Pegel des Testschalles (<i>sensation level</i>) in Abhängigkeit von der Zeit Δt nach dem Einschalten des Maskierers bzw. t_d nach dem Abschalten.	19
2.3.	Mithörschwellen resultierend aus Schmalbandrauschen von 60 dB mit den Mittenfrequenzen $f_c = 0,25$ kHz, 1 kHz und 4 kHz und den Bandbreiten $\Delta f_G = 100$ Hz, 160 Hz, und 700 Hz. Zusätzlich ist die Ruhehörschwelle gestrichelt eingezeichnet.	21
2.4.	Mithörschwellen maskiert durch einen 1 kHz Sinuston für verschiedene Testton-Pegel L_M	22
2.5.	Grafische Darstellung der Frequenzgruppen nach Zwicker über der Frequenz	23
2.6.	Einige Mithörschwellen maskiert durch mehrere bandbegrenzte Rauschsignale mit je 60 dB Schalldruckpegel, aufgetragen über die Frequenzgruppen	24
2.7.	Darstellung der Pegelabhängigkeit der oberen Flanke einer Mithörschwelle, erzeugt durch Schmalbandrauschen	25
3.1.	Schematische Darstellung des Frequenzgangs einer Analyse-Filterbank .	26
3.2.	Blockschaltbild einer Zweikanal Analyse-Filterbank	27
3.3.	Amplitudengänge der Analyseteilfilter einer Zweikanal Filterbank . . .	28
3.4.	Blockschaltbild einer Zweikanal Synthese-Filterbank	28
3.5.	Amplitudengänge der Syntheseteilfilter einer Zweikanal Filterbank . . .	29
3.6.	Verzerrungs- und Aliasfunktion einer perfekt rekonstruierenden Zweikanal Filterbank	31
3.7.	Filterbank in Baumstruktur	32
3.8.	Filterbank in Parallelstruktur	32
3.9.	Darstellung der äquidistanten Frequenzverschiebung eines Prototypen mittels Diskrete Fourier Transformation (DFT) Algorithmus	33

3.10. Blockschaltbild einer DFT-Analyse-Filterbank	34
3.11. Blockschaltbild einer DFT-Synthese-Filterbank	35
3.12. Prototypenfilter für eine M=4 DFT-Filterbank	36
3.13. Simulationsergebnisse einer M=4 DFT-Filterbank	37
3.14. Gleichmäßige Aufteilung des Frequenzbereichs durch Modulation des Prototypen	39
3.15. Blockschaltbild einer M-Kanal Cosinus-modulierten Analyse-Filterbank	40
3.16. Finales Blockschaltbild der Cosinus-modulierten Filterbank	42
3.17. Prototypenfilter für eine M=4 Cosinus-modulierte Filterbank	43
3.18. Simulationsergebnisse einer M=4 Cosinus-modulierten Filterbank . . .	44
3.19. Darstellung ausgewählter Teilbänder der Cosinus-modulierten Filter- bank für M=4, mit zugehörigen Aliaskomponenten äquivalent zum Blockschaltbild Abbildung 3.15	45
3.20. Optimiertes Prototypenfilter für eine M=4 Cosinus-modulierte Filterbank	47
3.21. Simulationsergebnisse einer M=4 Cosinus-modulierten Filterbank mit optimiertem Prototypen	48
4.1. Blockschaltbild eines MPEG-1 Layer I und II Encoders	50
4.2. Flussdiagramm des Entwurfs der MPEG-1 Analyse-Filterbank	52
4.3. Vergleich der in [Sta91] gegebenen Koeffizienten mit denen basierend auf der Methodik aus Abschnitt 3.6 hergeleiteten Koeffizienten	53
4.4. Darstellung des verwendeten Filterbank Prototypen	54
4.5. Blockschaltbild der Cosinus-modulierten Analyse-Filterbank mit Ab- wärtstastung	55
4.6. Simulationsergebnisse der M=32 Cosinus-modulierten Filterbank	56
4.7. Vergleich von Fensterfunktionen	58
4.8. Ablaufdiagramm des psychoakustischen Modell 1	60
4.9. Schematische Darstellung des Zusammenhangs von Indices k und m . Alle Funktionen in m beinhalten dabei eine zu hohen Frequenzen hin schrittweise dezimierte Darstellung als Funktionen in k	61
4.10. Schematische Darstellung der simultanen Maskierung	63
4.11. Für Simulationszwecke verwendetes Test-Signal	66
4.12. Maskierer und Mithörschwellen resultierend aus dem Test-Signal, auf- getragen auf der Frequenz-Skala	67
4.13. Maskierer und Mithörschwellen resultierend aus dem Test-Signal, auf- getragen auf der Bark-Skala	67
4.14. Globale Mithörschwelle $L_{T_g}(f)$ resultierend aus dem Test-Signal	68
4.15. Ermittlung des Signal- Mithörschwellenabstands aus Schalldruckpegel und minimaler Mithörschwelle	70
4.16. Ablaufdiagramm der dynamischen Bitzuweisung	71
4.17. Dynamische Bitzuweisung für das Test-Signal bei $v_B=192$ kbps	75
4.18. Kennlinie einer 3 Bit Quantisierung	77

4.19. MPEG-1 Audio Rahmen (grün: genutzter Inhalt)	77
4.20. Anordnung des Datenfelds	77
4.21. Blockschaltbild der Cosinus-modulierten Synthese-Filterbank mit Aufwärtstastung	78
4.22. Spektrogramme \mathcal{S}^{Swe} mit $v_B = 64$ kbps (Simulation)	80
4.23. Spektrogramme \mathcal{S}^{Swe} mit $v_B = 384$ kbps (Simulation)	80
4.24. Spektrogramme \mathcal{S}^{Bra} mit $v_B = 64$ kbps (Simulation)	81
4.25. Spektrogramme \mathcal{S}^{Bra} mit $v_B = 384$ kbps (Simulation)	81
4.26. Spektrogramme \mathcal{S}^{Joh} mit $v_B = 64$ kbps (Simulation)	82
4.27. Spektrogramme \mathcal{S}^{Joh} mit $v_B = 384$ kbps (Simulation)	82
4.28. Spektrogramme \mathcal{S}^{Kal} mit $v_B = 64$ kbps (Simulation)	83
4.29. Spektrogramme \mathcal{S}^{Kal} mit $v_B = 384$ kbps (Simulation)	83
5.1. Entwicklungsboard C6713 DSK der Firma <i>Spectrum Digital</i>	85
5.2. Blockschaltbild des TMS320C6713B	86
5.3. Darstellung des Serial Peripheral Interface (SPI)-Bus	87
5.4. Darstellung der Blockdauer von 8 ms bei unterschiedlichen Bitraten/-Kompressionsfaktoren	88
5.5. Blockschaltbild der DSK-Boards und Beschaltung der AIC23-Codecs mit dem Digitaler Signal Prozessor (DSP)	88
5.6. Flussdiagramm des Encoders unter Verwendung des Enhanced Direct Memory Access (EDMA)-Controllers. Der Encoder-Vorgang (farblich gekennzeichnet) wird im Abschnitt 5.4 detailliert dargestellt.	90
5.7. Messung des SPI Bus (Messdauer von zwei <i>FSX</i> Signalen) CH1: <i>FSX</i> , CH2: <i>CLKX</i> , CH3: <i>DX</i>	90
5.8. Messung des SPI Bus (Messdauer eines <i>FSX</i> Signals) CH1: <i>FSX</i> , CH2: <i>CLKX</i> , CH3: <i>DX</i>	91
5.9. DSK-Board mit aufgestecktem Adapterboard des IT-Labors an der HAW zur Messung eines Amplitudengangs mit 32 Bit Auflösung über I ² S	92
5.10. Amplitudengänge der gemessenen unteren 16 Teilbänder	93
5.11. Amplitudengänge der gemessenen oberen 16 Teilbänder	93
5.12. Gemessene Verzerrungsfunktion der Filterbank	94
5.13. Flussdiagramm des Encoders	95
5.14. Darstellung der 32-Bit Single-Präzision, [Tex06d, S. 71] s=Vorzeichen, e=Exponent, f=Mantisse	96
5.15. Differenzen unterschiedlicher Berechnungen der globalen Mithörschwelle $L_{T_g}(f)$	98
5.16. $X(k)$	99
5.17. $L_S(n)$	99
5.18. $L_{T_{min}}(n)$	99
5.19. $SMR(n)$	99
5.20. $BSPL(n)$ vor der Bitzuweisung	100

5.21. $BSPL(n)$ für $v_B = 192$ kbps	100
5.22. $MNR(n)$ vor der Bitzuweisung	100
5.23. $MNR(n)$ für $v_B = 192$ kbps	100
5.24. Flussdiagramm des Decoders	101
6.1. Amplitudengang DSK-Board	103
6.2. Amplitudengang bei $v_B = 384$ kbps	103
6.3. Amplitudengang bei $v_B = 192$ kbps	103
6.4. Amplitudengang bei $v_B = 128$ kbps	103
6.5. Amplitudengang bei $v_B = 96$ kbps	104
6.6. Amplitudengang bei $v_B = 64$ kbps	104
6.7. Darstellung des Aufbaus zur Messung der Spektrogramme	104
6.8. Spektrogramm \mathcal{S}^{Swe} mit $v_B = 64$ kbps (Messung)	105
6.9. Spektrogramm \mathcal{S}^{Swe} mit $v_B = 384$ kbps (Messung)	105
6.10. Spektrogramm \mathcal{S}^{Bra} mit $v_B = 64$ kbps (Messung)	106
6.11. Spektrogramm \mathcal{S}^{Bra} mit $v_B = 384$ kbps (Messung)	106
6.12. Spektrogramm \mathcal{S}^{Joh} mit $v_B = 64$ kbps (Messung)	107
6.13. Spektrogramm \mathcal{S}^{Joh} mit $v_B = 384$ kbps (Messung)	107
6.14. Spektrogramm \mathcal{S}^{Kal} mit $v_B = 64$ kbps (Messung)	108
6.15. Spektrogramm \mathcal{S}^{Kal} mit $v_B = 384$ kbps (Messung)	108
A.1. Darstellung der Polyphasen des Signals $x[n]$ für $M = 3$	114
A.2. Modulationsdarstellung für $M=3$	116
A.3. Spektren bei Dezimation für $M=3$	117
A.4. Spektren bei Interpolation für $M=3$	118
B.1. Spektrogramme \mathcal{S}^{Swe} bei allen Bitraten (Simulation)	119
B.2. Spektrogramme \mathcal{S}^{Bra} bei allen Bitraten (Simulation)	120
B.3. Spektrogramme \mathcal{S}^{Joh} bei allen Bitraten (Simulation)	121
B.4. Spektrogramme \mathcal{S}^{Kal} bei allen Bitraten (Simulation)	122
B.5. Spektrogramme \mathcal{S}^{Swe} bei allen Bitraten (Messung)	123
B.6. Spektrogramme \mathcal{S}^{Bra} bei allen Bitraten (Messung)	124
B.7. Spektrogramme \mathcal{S}^{Joh} bei allen Bitraten (Messung)	125
B.8. Spektrogramme \mathcal{S}^{Kal} bei allen Bitraten (Messung)	126

Abkürzungsverzeichnis

ADC	Analog-Digital-Converter
AKF	Autokorrelationsfunktion
CD	Compact Disc
CPU	Central Processing Unit
DAC	Digital-Analog-Converter
DCT	Diskrete Cosinus Transformation
DFT	Diskrete Fourier Transformation
DSP	Digitaler Signal Prozessor
EDMA	Enhanced Direct Memory Access
FFT	Fast Fourier Transformation
FLOPS	Floating Point Operations Per Second
ISO	International Organization for Standardization
LDS	Leistungsdichtespektrum
LUT	Lookup Tabelle
McBSP	Multichannel Buffered Serial Port
MPEG	Moving Picture Experts Group
MNR	Mask-to-Noise Ratio
PR	Perfekte Rekonstruktion
QMF	Quadrature Mirror Filter
SMR	Signal-to-Mask Ratio
SNR	Signal-to-Noise Ratio
SPI	Serial Peripheral Interface

1. Einleitung

In dieser Arbeit wird die Entwicklung eines Echtzeit-Systems zur Komprimierung von Audiosignalen bzw. zur Reduzierung der Bitrate auf Grundlage des MPEG-1 Verfahrens mittels Digitalen Signal Prozessors (DSP) vorgestellt. Dieses Kapitel dient zur Darstellung der Motivation sowie zur Betrachtung aktueller Verfahren. Darauf aufbauend wird die Aufgabenstellung formuliert. Das Kapitel schließt mit der Gliederung und einer Auflistung der in dieser Arbeit verwendeten Geräte.

1.1. Motivation

Mit dem rasanten Fortschritt der Informationstechnologie geht auch ein Anstieg der Datenmenge einher. Die Reduzierung der Daten ist dabei für viele Systeme meist unabwendbar. Dies gilt besonders für Video- und Audio-Signale. Das MPEG-1 Verfahren bietet eine Möglichkeit der Dezimierung von Audiodaten, welche grundlegend in zwei Anwendungsbereiche unterteilt werden kann:

Speichermedien bezogene Audiokomprimierung:

Dies ist eine aus dem Alltag bekannte Anwendung des MPEG-1 Audio Verfahrens. Werden beispielsweise eine große Menge Musikstücke im Rohformat auf einen Datenträger kopiert, so führt dieser Vorgang mit andauernder Anzahl an Musikstücken zur Auslastung des zur Verfügung stehenden Speichers. Ein Compact Disc (CD)-Musikstück besitzt üblicherweise eine Datenrate von 705,6 kbps (44,1 kHz Abtastrate und 16 Bit Auflösung). Ein drei Minuten Lied mit Stereo Klang hätte somit eine Dateigröße von

$$d = 2 \text{ CH} \cdot 16 \text{ Bit} \cdot 44,1 \text{ kHz} \cdot 3 \text{ Minuten} = 31,75 \text{ MByte.}$$

Es ist nicht unüblich, dass Kompressionsraten von 1:12 und höher erreicht werden können. Audiokompression dient somit zur Entlastung des Datenträgers. Dies hat in der Vergangenheit auch dazu geführt, dass Online-Handelsplattformen ihre Musikstücke bereits komprimiert zum Datei-Download anbieten. Dabei wird unter anderem das bekannte MP3 Format (MPEG-1 Layer III) als Standard genutzt.

Echtzeit Bitratenreduktion:

Die zweite Anwendung ist mehr passiv aus dem Alltag bekannt. Das VHF-Band III¹ ist der Frequenzbereich für DAB (*Digital Audio Broadcasting*)-Kanäle. Die spezifizierete Bandbreite eines Kanals liegt bei 1,536 MHz. Ohne Codierung und OFDM²-Technik fände hier genau ein Stereo Audiosignal mit 48 kHz Abtastrate und 16 Bit Auflösung Raum. Der Kanal 5C beispielsweise streamt jedoch ganze 18 Sender parallel. zu Anfangszeiten des DAB wurde das Moving Picture Experts Group (MPEG)-1 Layer II (MP2) Verfahren verwendet [MV13]. Heutzutage wird das MPEG-4 AAC Verfahren genutzt [UKW]. Durch die Echtzeit-Reduzierte Bitrate des Audiostreams wird somit wichtige Übertragungsbandbreite eingespart.

1.2. Stand der Technik

Die Entwicklung des ersten MPEG-Kompressionsverfahrens geht auf die frühen 80er Jahre zurück. Das bekannteste Verfahren dieses Standards ist der Layer III, besser bekannt als MP3, welches im November 1992 zugelassen wurde. Das Konzept beruht neben psychoakustischen Effekten auf einer Teilbandcodierung mittels Filterbank.

Eine Erweiterung wurde bereits zwei Jahre später mit dem MPEG-2 Verfahren vorgestellt. Dabei wurden zunächst andere Abtast- und Bitraten realisiert. Später kam eine Neuentwicklung mit dem Advanced Audio Coding (*AAC*) hinzu. Dieses Verfahren beruht nicht wie MP3 auf der Teilbandcodierung mittels Filterbank, sondern nutzt stattdessen eine modifizierte diskrete Cosinus-Transformation (*MDCT*). Erstmalig gelang hiermit eine Komprimierung von Mehrkanälen wie 5.1³ statt. Diese Methode fand kurzzeitig auf DVDs Anwendung. Abgelöst wurde sie 1995 durch Dolby Digital, welches durch das Adaptive Transform Coding 3 (*AC-3*) realisiert wird.

MPEG-3 hätte der Standard für High Definition Television (*HDTV*) werden sollen. Es genügte jedoch die Erweiterung von MPEG-2. MPEG-3 wurde daher nicht mehr verabschiedet [Sal06].

Niedrigere Datenraten werden mittlerweile durch das MPEG-4 High Efficiency Advanced Audio Coding (*HE-AAC*) Verfahren erzielt. Dies nutzt beispielsweise das Digital Audio Broadcasting Plus (*DAB+*) bei Datenraten von 48 kbps bis 128 kbps [UKW]. Dabei erzielt HE-AAC eine bessere Klangqualität als die Methode des MPEG-2 AAC [SPA07].

¹Das VHF-Band III ist die Bezeichnung eines Frequenzbereichs mit Ultrakurzwelle (VHF: *engl. Very High Frequency*) zw. 174 MHz und 230 MHz.

²OFDM steht für Orthogonal Frequency-Division Multiplexing und ist ein Vielträger-Verfahren, welches Verwendung in der Breitbandübertragung von digitalen Daten findet.

³5.1 bezeichnet ein Mehrkanal(-Umfeld)-Tonsystem (Surround-Sound), bei dem drei Hauptkanäle (Vorne-Links, -Center, -Rechts), zwei Umfeldkanäle (Hinten-Links und -Rechts) sowie ein separater Tieftonkanal zur Verfügung steht.

1.3. Aufgabenstellung

In dieser Arbeit soll eine Echtzeit-Audiokompression auf einem digitalen Signalprozessor realisiert werden. Dabei besteht die Aufgabe darin, die Bitrate eines beliebigen Audiosignal durch einen Encoder-DSP zu reduzieren und durch einen Decoder-DSP akustisch wiederzugeben. Bei dem Eingangsaudiosignal handelt es sich bspw. um ein Musikstück in CD-Qualität (Abtastfrequenz 48 kHz oder 44,1 kHz mit 16 Bit Auflösung).

Echtzeit bedeutet hierbei, dass der Encoder-DSP stets zur Verarbeitung des anfallenden Audiostreams bereit ist. Die resultierenden bitreduzierten-Daten werden nicht auf einem Medium gespeichert, sondern an einen Decoder-DSP übertragen. Dieser fügt die Daten wieder zusammen und gibt das rekonstruierte Signal über einen Lautsprecher aus.

Die gewünschte Bitrate bzw. der Kompressionsfaktor ist dabei beliebig einstellbar. Abbildung 1.1 zeigt das Prinzipschaltbild. Als Hardwareplattform steht der TMS320C6713 DSP der Firma *Texas Instruments* zur Verfügung.

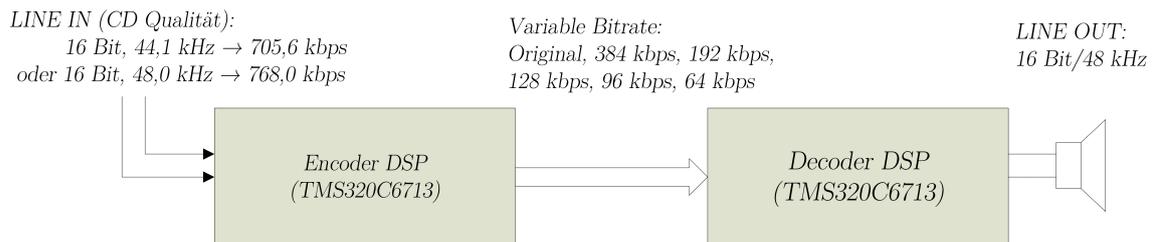


Abbildung 1.1.: Prinzipschaltbild des Gesamtsystems

Bei dem Verfahren zur Bitratenreduktion wird sich weitestgehend an die verlustbehaftete Audiokompression MPEG-1 Layer I gehalten, welches der Standard ISO/IEC 11172-3 [Sta91] beschreibt. Hierbei kommt ein psychoakustisches Modell zum Einsatz, welches nur Signalanteile isoliert, die vom menschlichem Gehör wahrgenommen werden können. Das Gesamtsystem wird zunächst in einer *MATLAB*-Simulation entwickelt und untersucht. Anschließend erfolgt die Implementierung auf der Hardwareplattform.

Abschließende Messungen von Kurzzeit-Spektrogrammen und Amplitudengängen sollen zur Beurteilung des Verfahrens dienen. Des Weiteren soll ein abschließender Hörtest für eine subjektive Beurteilung der Kompression bzw. der Audioqualität dienen.

1.4. Gliederung

Kapitel 2 gibt eine Einführung in die Grundlagen der Psychoakustik, auf denen das MPEG-1 Verfahren beruht.

Kapitel 3 befasst sich mit der Zerlegung von Audiosignalen in mehrere Teilbänder mittels Filterbänken. Zunächst wird in Abschnitt 3.1 eine Zweikanal-Filterbank beschrieben. Über die DFT-Filterbank in Abschnitt 3.5 wird schließlich die in Abschnitt 3.6 beschriebene Cosinus-modulierte-Filterbank hergeleitet, welche Anwendung beim MPEG-1 Verfahren findet.

Kapitel 4 thematisiert jeden Schritt des MPEG-1 Verfahrens und untersucht diese anhand von Simulationen. Der Decoder zur Signalsynthese wird in Abschnitt 4.8 beschrieben.

Die verwendete Hardware und die Implementierung des Systems auf dem DSP wird in Kapitel 5 dargestellt.

Eine abschließende Evaluation anhand von Messungen und Beurteilungen von Testhörern befindet sich in Kapitel 6.

Kapitel 7 fasst die Ergebnisse zusammen und gibt einen Ausblick auf mögliche weitere Arbeiten.

1.5. Notation

Um die Lesbarkeit dieser Arbeit zu erhöhen, wird eine einheitliche Notation verwendet. Diese ist in Tabelle 1.1 zusammengefasst.

Bezeichnung	Kennzeichnung	Beispiel
Skalar	kleines Zeichen	a
Spalten Vektor	kleines, fettes Zeichen	\mathbf{a}
Matrix	großes, fettes Zeichen	\mathbf{A}
transponierte(r) Matrix / Vektor	hochgestellter Buchstabe T	$\mathbf{a}^T, \mathbf{A}^T$
konjugiert komplexe(r) Matrix / Vektor	hochgestellter Stern	$\mathbf{a}^*, \mathbf{A}^*$
Zeitdiskrete Signalfolge	eckige Klammern	$a[n] = a(nT)$
Mittelwert	Strich über Symbol	\bar{a}
Rekonstruktion	Dach über Symbol	\hat{x}

Tabelle 1.1.: Notation

Des Weiteren sei an dieser Stelle erwähnt, dass Abbildungen im Frequenzbereich stets mit einer auf π normierten ω -Achse dargestellt werden, wobei ω die auf die Abtastfrequenz normierte Kreisfrequenz darstellt.

1.6. Geräteliste

In Tabelle 1.2 sind die verwendeten Geräte aufgelistet.

Nr.	Gerätebezeichnung	Zweck/Verwendung
1.	TI TMS320C6713 DSP Starter Kit (DSK)	DSP Entwicklungsboard
2.	Rohde & Schwarz UPV Audio Analyzer	Digitale und analoge Audiomessungen
3.	Tektronix MSO 2000B	Oszilloskop
4.	PeakTech 3500FG	Signalgenerator
5.	Technics RP-DJ1210E-S	Kopfhörer für Hörtest
6.	Wavemaster 1520	Lautsprecher/Verstärker für Hörtest

Tabelle 1.2.: Verwendete Geräte

2. Psychoakustik

Eine auf das menschliche Gehör angepasste Codierung wie das MPEG-1 Verfahren basiert auf psychoakustischen Untersuchungen von E. Zwicker mit der Hilfe von gut geübten Testhörern [ZF99]. Die für die Codierung gewonnenen Erkenntnisse werden in diesem Kapitel weitestgehend vorgestellt, da sie die Grundlage für folgende Kapitel darstellen.

Das Gehör analysiert Schallereignisse in Frequenzgruppen. Ziel ist es, Audiosignale in diese gehörgerechten Frequenzgruppen zu zerlegen und die resultierenden Teilbandsignale zu codieren. Dazu wird die Ruhehörschwelle und Maskierungseffekte ausgenutzt, die folgend erläutert werden. Zunächst bedarf es jedoch einiger Definitionen. In der Psychoakustik ist der Schalldruckpegel

$$L_S/\text{dB} = 20 \cdot \log_{10}(p/p_0) \quad (2.1)$$

zur Beschreibung der Stärke eines Schallereignisses gebräuchlich, da die Empfindung des menschlichen Gehörs in etwa logarithmisch verläuft. Dabei ist p der Schalldruck und p_0 die Bezugsgröße, welche auf

$$p_0 = 2 \cdot 10^{-5} \text{ Pa} \quad (2.2)$$

festgelegt wurde [ZF99]. Des Weiteren sind in der Psychoakustik Töne bzw. tonale Schallsignale gebräuchlich. Einen tonalen Charakter haben Schallsignale, die sich zeitlich periodisch wiederholen, wie z.B. der Schall von schwingenden Saiten oder eines Sinussignals.

2.1. Absolute Hörschwelle

Die absolute Hörschwelle, welche auch als Ruhehörschwelle bezeichnet wird, ist derjenige Schalldruckpegel eines Sinustons, der in Abhängigkeit von seiner Frequenz gerade noch wahrgenommen wird. Die absolute Hörschwelle ist durch die Physiologie des menschlichen Gehörs gegeben. Sie unterscheidet sich von Mensch zu Mensch und maßgeblich mit zunehmendem Alter. Daher kann keine allgemeingültige Funktion definiert werden. Es wurden jedoch Messverfahren entwickelt, um die Hörschwelle (und

auch verschiedene Mithörschwellen) einer beliebigen Testperson zu ermitteln. Häufig wird hierfür die Methode des pendelnden Einregelns⁴ angewandt [ZF99]. Durch die Mittelung einer hohen Anzahl an Messungen kann eine Näherungsfunktion gebildet werden. Nach E. Terhardt [ZÖ8] gilt für die absolute Hörschwelle

$$L_{T_q}/\text{dB} = 3.64 \cdot (f/\text{kHz})^{-0.8} - 6.5 \cdot \exp\left(-0.6 \cdot (f/\text{kHz} - 3.3)^2\right) + 10^{-3} \cdot (f/\text{kHz})^4, \quad (2.3)$$

welche in Abbildung 2.1 dargestellt ist.

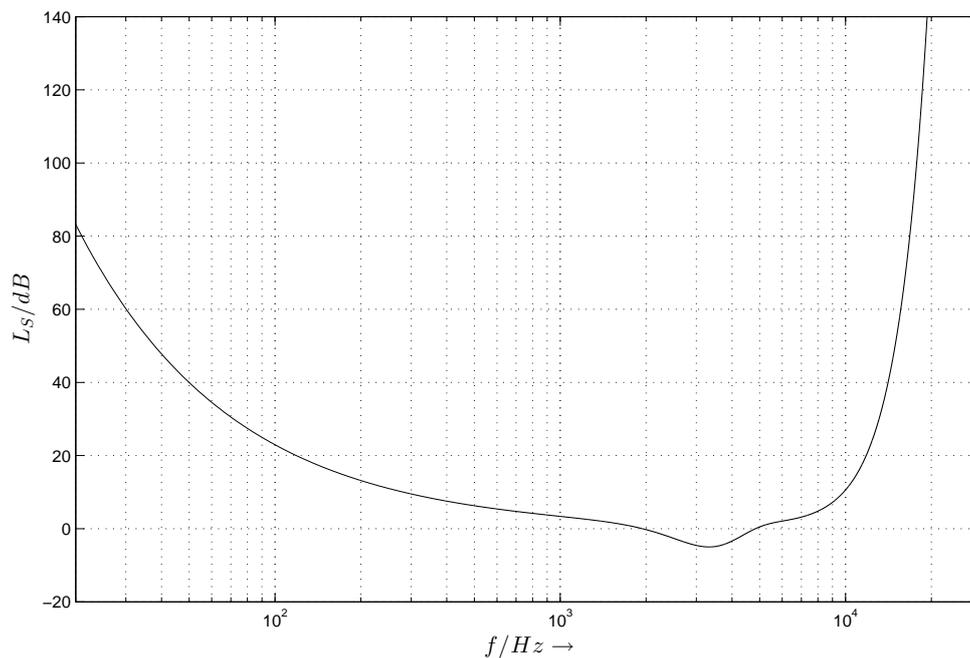


Abbildung 2.1.: Absolute Hörschwelle nach E. Terhardt

Bei tiefen Frequenzen liegt die Ruhehörschwelle verhältnismäßig hoch. Sie erreicht bei 50 Hz einen Pegel von etwa 40 dB und bei 300 Hz einen Pegel von etwa 15 dB. Zwischen 2 kHz und 5 kHz ist das Gehör am empfindlichsten und damit auch am anfälligsten für Schädigungen. Oberhalb von ca. 12 kHz steigt die Ruhehörschwelle zügig an, was dazu führt, dass Hörempfindungen von Tönen auch bei sehr starken Pegeln wenig bis gar nicht mehr hervorgerufen werden.

⁴Die Methode des pendelnden Einregelns (oder auch Békésy-Methode) ist ein Verfahren, welches angewandt wird, um Hörschwellen bei Versuchspersonen zu bestimmen.

2.2. Maskierung und Mithörschwelle

Befinden sich Töne oberhalb der absoluten Hörschwelle und können dennoch vom Gehör nicht wahrgenommen werden, handelt es sich um den Maskierungseffekt. Dieser Effekt ist aus dem Alltag bekannt. Für eine Unterhaltung unter zwei Personen ist im allgemeinen keine große Lautstärke notwendig. Tritt jedoch plötzlich ein starkes Störgeräusch auf, wie bspw. ein schneller Zug, der sehr nah an den Personen vorbeifährt, kann es dazu kommen, dass der Sprecher unhörbar wird. In diesem Moment wird der Sprecher übertönt bzw. maskiert. Man spricht auch von einer Verdeckung. Das Geräusch des Zugs wird in diesem Fall auch als Maskierer bezeichnet. Wird der Sprachpegel soweit angehoben, dass er erneut hörbar ist, so wurde die sogenannte Mithörschwelle überschritten und die Maskierung ist aufgehoben. Der Maskierungseffekt ist ein wichtiger Teil in der Psychoakustik und folglich auch für die Audiokompression. Man unterscheidet zwischen Zeitbereichs- und Frequenzbereichsmaskierung.

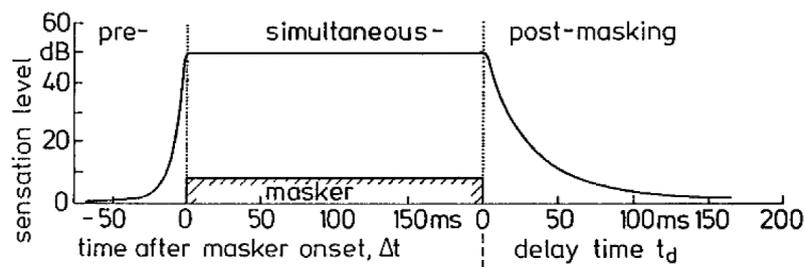


Abbildung 2.2.: Verdeckungs- bzw. Maskierungseigenschaft des menschlichen Gehörs. Nicht-simultane Vormaskierung (*engl. pre-masking*), nicht-simultane Nachmaskierung (*engl. post-masking*) und simultane Maskierung. Aufgetragen ist der Pegel des Testschalls (*sensation level*) in Abhängigkeit von der Zeit Δt nach dem Einschalten des Maskierers bzw. t_d nach dem Abschalten [ZF99].

2.2.1. Zeitbereichsmaskierung

Bei der Zeitbereichsmaskierung treten Schalle bzw. Geräusche, wie im obigen Beispiel der Unterhaltung zweier Personen, nicht gleichzeitig auf. Es handelt sich somit um eine nicht-simultane Maskierung, welche in zwei Varianten unterschieden wird. Tritt ein Testschall zeitlich vor einem Störschall bzw. Maskierer auf, so handelt es sich um eine Vormaskierung. Diese wirkt über eine Zeit von $\Delta t \approx -20 \text{ ms}$. Im umgekehrten Fall tritt eine Nachmaskierung auf, welche einen dominanteren Effekt mit $t_d \approx 100 \text{ ms}$ hat. Die zeitlichen Grenzen zur Hörbarkeit werden als Nach- bzw. Vorhörschwelle bezeichnet [ZF99]. Abbildung 2.2 verdeutlicht das Schema des Maskierungseffekts.

Die Zeitbereichsmaskierung ist kein Gegenstand im MPEG-1 Layer I Standard. Hier wird stattdessen der Effekt der Frequenzbereichsmaskierung ausgenutzt. Beide Maskierungseffekte finden beispielsweise im Layer III (MP3) des MPEG-1 Verfahren Anwendung [Sta91].

2.2.2. Frequenzbereichsmaskierung

Treten Testschall und Störschall zeitgleich auf, handelt es sich um eine simultane Maskierung. Was bei der Zeitbereichsmaskierung die Grenzen der Hörbarkeit (die Nach- und Vorhörschwelle) darstellt, ist bei der Maskierung im Frequenzbereich die Mithörschwelle. Sie gibt den Schalldruckpegel eines Testschalls an, den dieser haben muss, damit er neben dem Störschall gerade noch wahrgenommen bzw. mitgehört werden kann. Das bedeutet auch, dass die Mithörschwelle immer über oder auf der Ruhehörschwelle liegt. Wie die Zeitbereichsmaskierung kann auch die Frequenzbereichsmaskierung in zwei Varianten unterschieden werden.

Mithörschwelle für maskierendes bandbegrenztes Rauschen

Abbildung 2.3 zeigt die Mithörschwellen von drei Störschallen in Form von bandbegrenztem Rauschen mit je einem Schalldruckpegel von 60 dB. Auffällig ist, dass die Mithörschwelle bei 0,25 kHz breiter ausfällt als die beiden sich ähnelnden bei 1 kHz und 4 kHz. Dabei ist die Bandbreite des niederfrequenten Maskierungsrauschens geringer. Mit tiefer werdenden Frequenzen setzt offenbar eine Veränderung der Mithörschwellen-Form ein. Das Maximum der Mithörschwelle tritt an der Maskierer-Mittenfrequenz auf. Mit steigender Frequenz scheint das Maximum abzunehmen, wobei die Steilheit der Schwelle zunimmt.

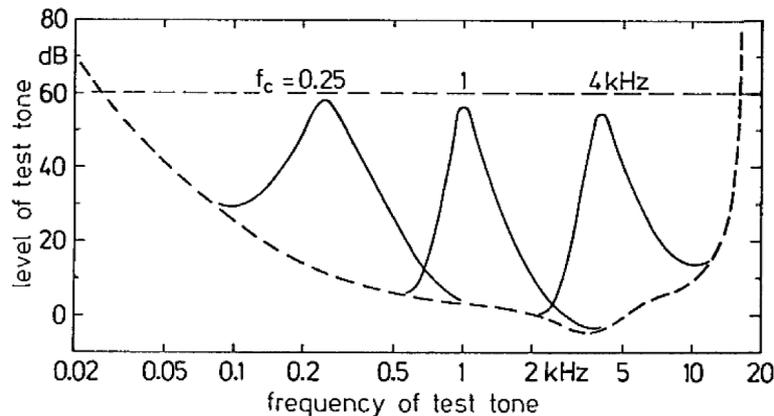


Abbildung 2.3.: Mithörschwellen resultierend durch Schmalbandrauschen von 60 dB mit den Mittenfrequenzen $f_c = 0,25$ kHz, 1 kHz und 4 kHz und den Bandbreiten $\Delta f_G = 100$ Hz, 160 Hz, und 700 Hz. Zusätzlich ist die Ruhehörschwelle gestrichelt eingezeichnet [ZF99].

Mithörschwelle für maskierende Sinustöne

Die Bestimmung der Mithörschwelle, welche durch einen einzigen Sinuston entsteht, kann nicht so eindeutig ermittelt werden, wie die eines rauschartigen Tons. Zwicker beschreibt in seinen Untersuchungen, dass die Testhörer Schwebungen in der Umgebung von Sinustönen bei $f = 1$ kHz, 2 kHz und 3 kHz wahrnahmen [ZF99]. Ein weiterer Effekt entstand, wenn zusätzlich zu einem Testton von bspw. $f_1 = 1$ kHz ein weiterer Ton mit $f_2 = 1,4$ kHz eingeblendet wurde. Bei geringen Pegeln des zweiten Tons wurde ein Differenzton (bei $f_d = 600$ Hz) wahrgenommen. Erst mit steigendem Pegel wurde der eingeblendete Testton erkannt. Der Differenzton wird durch nichtlineare Verzerrungen im Gehör erzeugt. Unter der Verwendung von apparativen Mitteln, welche die Differenztöne vermindern, kann die Mithörschwelle einzelner Töne gemessen oder wenigstens geschätzt werden. So zeigt Abbildung 2.4 die Mithörschwelle maskiert durch einen 1 kHz Sinuston mit verschiedenen Test-Pegeln. Für andere Maskierungsfrequenzen verhalten sich die Mithörschwellen in entsprechender Form. Dabei gilt auch bei Sinustönen: Bei tiefen Frequenzen der Maskierer werden die Mithörschwellen als Funktion der Testtonfrequenz etwas breiter. Bei Frequenzen oberhalb von 500 Hz treten kaum Veränderungen auf [ZF99].

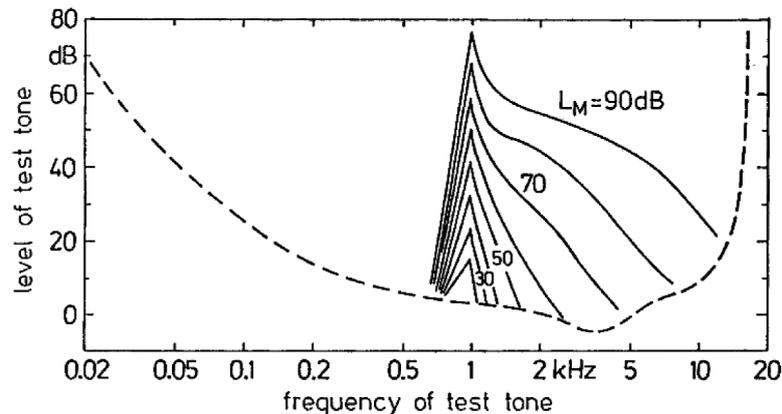


Abbildung 2.4.: Mithörschwellen maskiert durch einen 1 kHz Sinuston für verschiedene Testton-Pegel L_M [ZF99].

Die vorgestellten Effekte zeigen auf, dass in Audioschalle bzw. -signalen Irrelevanzen vorhanden sind. Diese nutzt die Audiokompression um eine Datenreduktion durchführen zu können.

2.3. Frequenzgruppen

Wie am Anfang des Kapitels bereits angedeutet, analysiert das menschliche Gehör Schalle in Frequenzgruppen (*engl. critical-band rate*). Diese spielen nicht nur bei der Empfindung der Tonhöhe eine wesentliche Rolle, sondern auch bei der Lautstärke- und Richtungsempfindung. Ein analytischer Ausdruck für die Transformation der linearen Frequenzskala in eine gehörrichtige Skala ist gegeben durch [ZF99]:

$$z/\text{Bark} = 13 \cdot \arctan\left(0,76 \frac{f}{\text{kHz}}\right) + 3,5 \cdot \arctan\left(\frac{f}{7,5 \text{ kHz}}\right)^2 \quad (2.4)$$

Dabei ist z die Tonheit mit der Einheit Bark. Die einzelnen Frequenzgruppen besitzen die Bandbreite (*engl. critical bandwidth*)

$$\Delta f_G/\text{Hz} = 25 + 75 \left(1 + 1,4 \left(\frac{f}{\text{kHz}}\right)^2\right)^{0,69} \quad (2.5)$$

und sind für den Frequenzbereich von $f=0$ Hz...15,5 kHz in Tabelle 2.1 aufgelistet. Abbildung 2.5 stellt diese grafisch dar. Ganz deutlich zu erkennen ist, dass das Gehör in den tieferen Frequenzen eine feinere Auflösung besitzt als in den hohen Frequenzen.

z/Bark	f_u/Hz	f_o/Hz	$\Delta f_G/\text{Hz}$	f_m/Hz
0	0	100	100	50
1	100	200	100	150
2	200	300	100	250
3	300	400	100	350
4	400	510	110	450
5	510	630	120	570
6	630	770	140	700
7	770	920	150	840
8	920	1080	160	1000
9	1080	1270	190	1170
10	1270	1480	210	1370
11	1480	1720	240	1600
12	1720	2000	280	1850
13	2000	2320	320	2150
14	2320	2700	380	2500
15	2700	3150	450	2900
16	3150	3700	550	3400
17	3700	4400	700	4000
18	4400	5300	900	4800
19	5300	6400	1100	5800
20	6400	7700	1300	7000
21	7700	9500	1800	8500
22	9500	12000	2500	10500
23	12000	15500	3500	13500
24	15500			

Tabelle 2.1.: Frequenzgruppen nach Zwicker

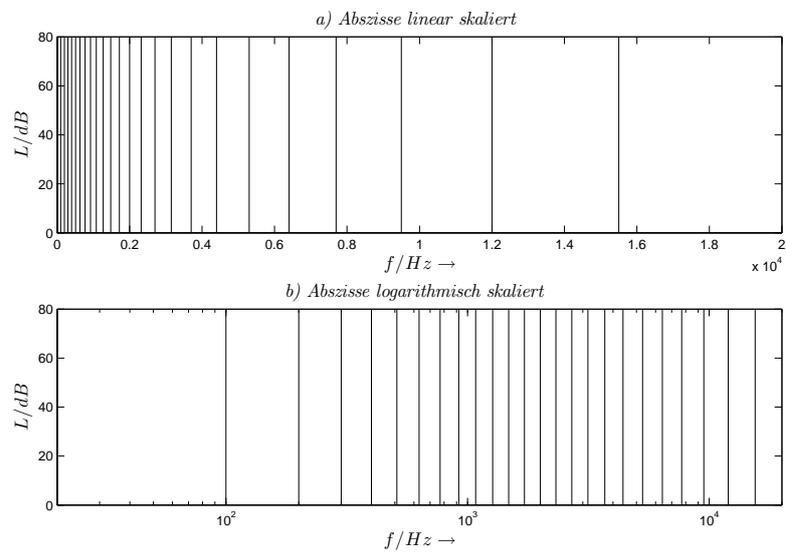


Abbildung 2.5.: Grafische Darstellung der Frequenzgruppen nach Zwicker über der Frequenz

Trägt man einige Mithörschwellen maskiert durch bandbegrenztetes Rauschen über die Frequenzgruppen auf, wird deutlich, dass die Transformation in die Tonheit einen Vorteil mit sich bringt, vgl. Abbildung 2.6.

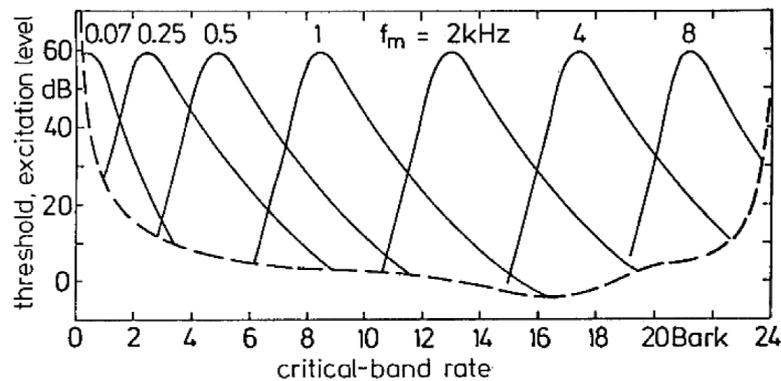


Abbildung 2.6.: Einige Mithörschwellen maskiert durch mehrere bandbegrenztetes Rauschsignale mit je 60 dB Schalldruckpegel, aufgetragen über die Frequenzgruppen [ZF99]

Für gleiche Erregungs-Schalldruckpegel sind die Muster im Wesentlichen durch Verschiebung in horizontaler Richtung identisch. Lediglich für sehr tieffrequentes Schmalbandrauschen ist die obere Flankenerregung etwas steiler. Die untere Flanke ist pegelelunabhängig und besitzt die Steigung

$$S_1 = 27 \text{ dB/Bark.} \quad (2.6)$$

Die obere Flanke besitzt mit

$$S_2 = 24 + 0,23 \left(\frac{f_{mi}}{\text{kHz}} \right)^{-1} - 0,2 \cdot \frac{L_G(i)}{\text{dB}} \text{ dB/Bark} \quad (2.7)$$

bei unterschiedlichen Erreger-Pegel L_G eine Abhängigkeit, vgl. Abbildung 2.7.

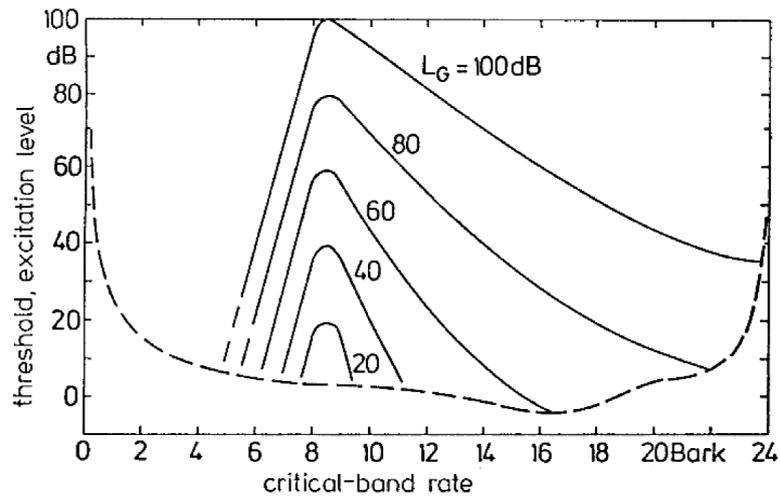


Abbildung 2.7.: Darstellung der Pegelabhängigkeit der oberen Flanke einer Mithörschwelle, erzeugt durch Schmalbandrauschen [ZF99]

Aufbauend auf die Ergebnisse von Zwicker sind einige mathematische Modelle für Maskierer bzw. Mithörschwellen aufgestellt worden, welche zusammengefasst in [ZÖ8] zu finden sind. Auf die Beschreibung der Modelle wird an dieser Stelle verzichtet, da das Modell, welches der MPEG-1 ISO Standard verwendet, in Kapitel 4 vorgestellt wird.

3. Filterbank

Eine Filterbank führt eine spektrale Zerlegung eines Signals in eine beliebige Anzahl an Frequenzbänder durch. Es wird dabei auch von der Signal-Analyse bzw. der Analyse-Filterbank gesprochen. Die nachgeschaltete Signal-Synthese bzw. Synthese-Filterbank setzt die Frequenzbänder wieder zusammen. Dabei können im gewissen Maße Rekonstruktionsfehler entstehen. Wird das ursprüngliche Signal unverfälscht rekonstruiert, ist von einer perfekten Rekonstruktion die Rede. Filterbänke bestehen mindestens aus einem Tiefpass und einem Hochpass. Je nachdem in wie viele Frequenzbänder zerlegt wird, sind ein oder mehrere Bandpässe enthalten, vgl. Abbildung 3.1.

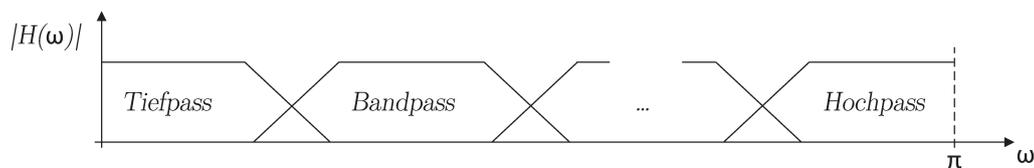


Abbildung 3.1.: Schematische Darstellung des Frequenzgangs einer Analyse-Filterbank

Die am häufigsten verwendete Anwendung von Filterbänken ist die Teilbandcodierung (*engl. Subband coding*), welche ein wichtigen Teil des MPEG-1 Verfahrens darstellt. Dies gilt für die Audio- sowie für die Bilddatenkompression [Fli93].

Dieses Kapitel wird zunächst durch die Zweikanal Filterbank eingeleitet (Abschnitt 3.1), danach werden die gängigsten M-Kanal Filterbankstrukturen vorgestellt (Abschnitt 3.4), bevor schließlich die für das MPEG-1 Verfahren genutzte Filterbank hergeleitet wird (Abschnitt 3.5 und 3.6). Dabei werden einige mathematische Grundlagen verwendet, die dem Anhang A zu entnehmen sind.

3.1. Zweikanal Filterbank

Die Zweikanal Filterbank bietet die einfachste Zerlegung eines Signals durch einen Tiefpass $H_0(z)$ und einen Hochpass $H_1(z)$.

3.1.1. Analyse-Filterbank

Der Nyquist-Frequenzbereich von $0 \leq \omega \leq \pi$ (normiert auf die Abtastfrequenz) wird in zwei Hälften aufgeteilt. Beide gefilterten Signale besitzen näherungsweise eine Bandbreite $BW = \pi/2$ und können daher in der Abtastrate um den Faktor 2 dezimiert werden [Fli93]. Da aus praktischen Gründen keine idealen Filter realisierbar sind, wird dabei das Nyquist-Theorem verletzt und es muss mit Aliasing gerechnet werden. Eine spätere Interpolation kann jedoch dafür sorgen, dass das gesamte System das Nyquist-Theorem einhält. Abbildung 3.2 zeigt das Blockschaltbild einer Zweikanal Analyse-Filterbank.

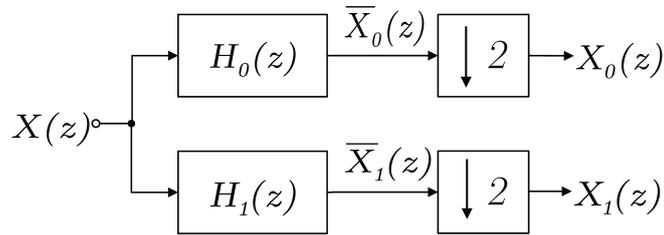


Abbildung 3.2.: Blockschaltbild einer Zweikanal Analyse-Filterbank

Die beiden Teilbänder vor der Abtastratenreduzierung lauten

$$\bar{X}_0(z) = X(z) \cdot H_0(z) \quad (3.1)$$

für den Tiefpassausgang sowie

$$\bar{X}_1(z) = X(z) \cdot H_1(z) \quad (3.2)$$

für den Hochpassausgang. Mit den im Anhang A definierten mathematischen Zusammenhängen folgt für die um den Faktor 2 dezimierten Teilbandsignale

$$X_0(z) = \frac{1}{2} \left(X(z^{1/2})H_0(z^{1/2}) + X(-z^{1/2})H_0(-z^{1/2}) \right) \quad (3.3)$$

und

$$X_1(z) = \frac{1}{2} \left(X(z^{1/2})H_1(z^{1/2}) + X(-z^{1/2})H_1(-z^{1/2}) \right). \quad (3.4)$$

Bzw. in Matrixschreibweise

$$\begin{aligned} \mathbf{x}(z) &= \frac{1}{2} \begin{bmatrix} H_0(z^{1/2}) & H_0(-z^{1/2}) \\ H_1(z^{1/2}) & H_1(-z^{1/2}) \end{bmatrix} \cdot \begin{bmatrix} X(z^{1/2}) \\ X(-z^{1/2}) \end{bmatrix} \\ &= \frac{1}{2} \cdot [\mathbf{H}^{(m)}(z^{1/2})]^T \cdot \mathbf{x}^{(m)}(z^{1/2}), \end{aligned} \quad (3.5)$$

wobei

$$\mathbf{x}(z) = \begin{bmatrix} X_0(z) & X_1(z) \end{bmatrix}^T \quad (3.6)$$

den Vektor der Teilbandsignale, $\mathbf{H}^{(m)}(z)$ die Modulationsmatrix der Analyse-Filterbank und $\mathbf{x}^{(m)}(z)$ den Modulationsvektor des Eingangssignals darstellen. Abbildung 3.3 zeigt beispielhaft die Amplitudengänge einer Zweikanal Analyse-Filterbank. Hier handelt es sich im Speziellen um *Quadrature-Mirror-Filter (QMF-Filter)*, da der Hochpass ein um $\pi/4$ abgebildetes Spiegelbild des Tiefpasses darstellt. Der Entwurfsprozess ist aus [Fli93, S. 187-Bsp. 6.2] entnommen.

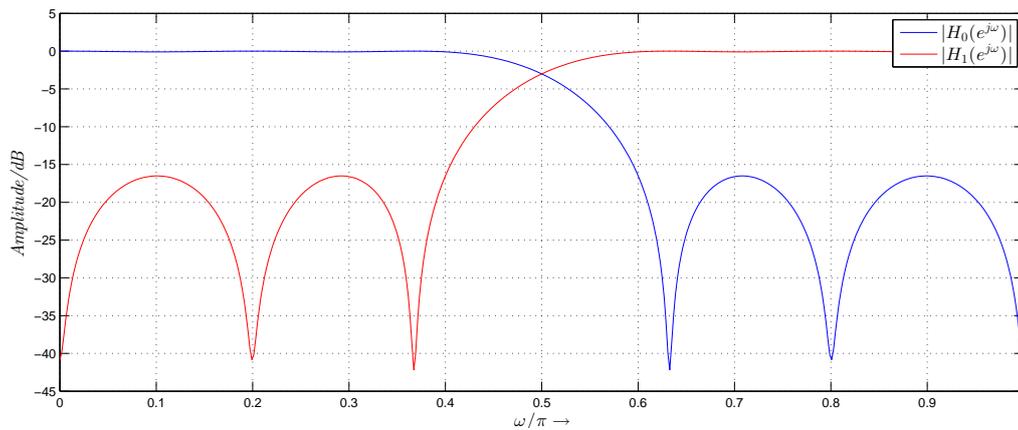


Abbildung 3.3.: Amplitudengänge der Analyseteilfilter einer Zweikanal Filterbank

3.1.2. Synthese-Filterbank

Das rekonstruierte Signal ist das Ergebnis der Synthese-Filterbank, welche ebenso aus zwei Filtern, $G_0(z)$ und $G_1(z)$ besteht. Bei $G_0(z)$ handelt es sich ebenso um einen Tiefpass, da dieser Pfad mit dem dezimierten Ausgang der Filterantwort des Tiefpasses $H_0(z)$ verbunden ist. Somit handelt es sich bei $G_1(z)$ um einen Hochpass. Abbildung 3.4 zeigt das Blockschaltbild einer Zweikanal Synthese-Filterbank.

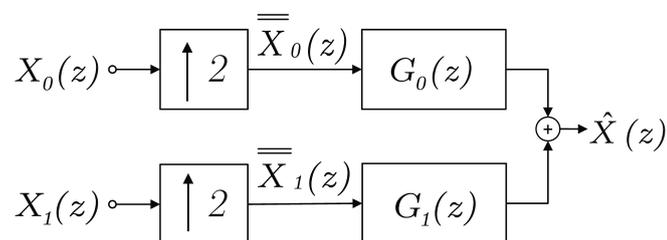


Abbildung 3.4.: Blockschaltbild einer Zweikanal Synthese-Filterbank

Das zusammengesetzte Signal entsteht aus der Summe der gefilterten, aufwärtsgetasteten Signale und wird durch die Rekonstruktionsgleichung

$$\hat{X}(z) = G_0(z)X_0(z^2) + G_1(z)X_1(z^2) \quad (3.7)$$

bzw. in Matrixschreibweise

$$\begin{aligned} \hat{X}(z) &= \begin{bmatrix} G_0(z) & G_1(z) \end{bmatrix} \cdot \begin{bmatrix} X_0(z^2) \\ X_1(z^2) \end{bmatrix} \\ &= \mathbf{g}^T(z) \cdot \mathbf{x}(z^2) \end{aligned} \quad (3.8)$$

beschrieben, wobei $\mathbf{g}(z)$ den Modulationsvektor der Synthesebank darstellt. Setzt man die Teilbandsignale aus Gleichung 3.3 und Gleichung 3.4 in Gleichung 3.7 ein, so erhält man das rekonstruierte Signal in Matrixschreibweise zu

$$\hat{X}(z) = \mathbf{g}^T(z) \cdot \frac{1}{2} \cdot [\mathbf{H}^{(m)}]^T \cdot \mathbf{x}(z) \quad (3.9)$$

bzw. ausgeschrieben zu

$$\begin{aligned} \hat{X}(z) &= \frac{X(z)}{2} (G_0(z)H_0(z) + G_1(z)H_1(z)) + \frac{X(-z)}{2} (G_0(z)H_0(-z) + G_1(z)H_1(-z)) \\ &= X(z)F_0(z) + X(-z)F_1(z). \end{aligned} \quad (3.10)$$

Abbildung 3.5 zeigt beispielhaft die Amplitudengänge der äquivalenten Zweikanal Synthese-Filterbank zur Analyse-Filterbank aus Abbildung 3.3

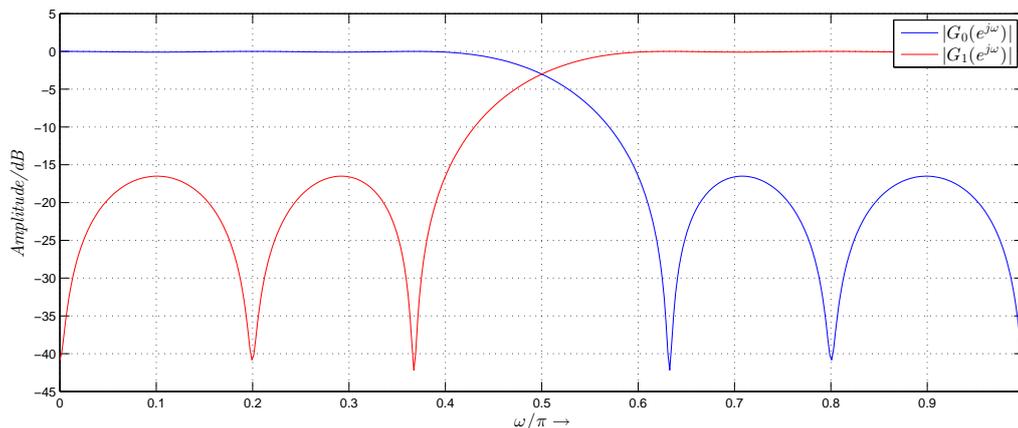


Abbildung 3.5.: Amplitudengänge der Syntheseteilfilter einer Zweikanal Filterbank

3.2. Verzerrungs- und Aliasfunktion

Aus Gleichung 3.10 kann die Funktion

$$F_0(z) = \frac{1}{2}G_0(z)H_0(z) + \frac{1}{2}G_1(z)H_1(z) \quad (3.11)$$

entnommen werden, welche das Übertragungsverhalten der Filterbank angibt. Allgemein wird auch von der Verzerrungsfunktion

$$F_{dist}(z) = \frac{1}{M} \sum_{i=0}^{M-1} |G_i(z)H_i(z)| \quad (3.12)$$

gesprochen. Diese kennzeichnet die Güte der Rekonstruktion. Das Aliasverhalten zeigt

$$F_1(z) = \frac{1}{2}G_0(z)H_0(-z) + \frac{1}{2}G_1(z)H_1(-z), \quad (3.13)$$

oder allgemein

$$F_{alias}(z) = \left(\sum_{l=1}^{M-1} \left| \frac{1}{M} \sum_{i=0}^{M-1} G_i(z)H_i(zW_M^l) \right|^2 \right)^{1/2}, \quad (3.14)$$

wobei W_M den Modulationsoperator (engl. twiddle factor) darstellt.

3.3. Perfekte Rekonstruktion

Von einer perfekten Rekonstruktion (PR) ist die Rede, wenn sich das rekonstruierte Signal lediglich durch eine Verzögerung und einer Skalaren-Gewichtung von dem Eingangssignal unterscheidet.

$$\hat{X}(z) = c \cdot z^{-p} \cdot X(z), \text{ wobei } c \neq 0 \text{ ein Skalar ist und } p \text{ ganzzahlig} \quad (3.15)$$

Die Filterbank-Übertragungsfunktion aus Gleichung 3.10 würde die PR-Bedingung erfüllen, sobald die Verzerrungsfunktion $F_0(z)$ eine Verzögerung darstellt und die Aliasfunktion $F_1(z)$ sich aufhebt. Abbildung 3.6 a) zeigt die aus der aufgezeigten Analyse- und Synthesefilterbank (Abbildungen 3.3 und 3.5) resultierende Verzerrungsfunktion der Zweikanal Filterbank. Abbildung 3.6 b) zeigt die Aliasfunktion. Abgesehen einiger Rechenungenauigkeiten besteht eine perfekte Rekonstruktion.

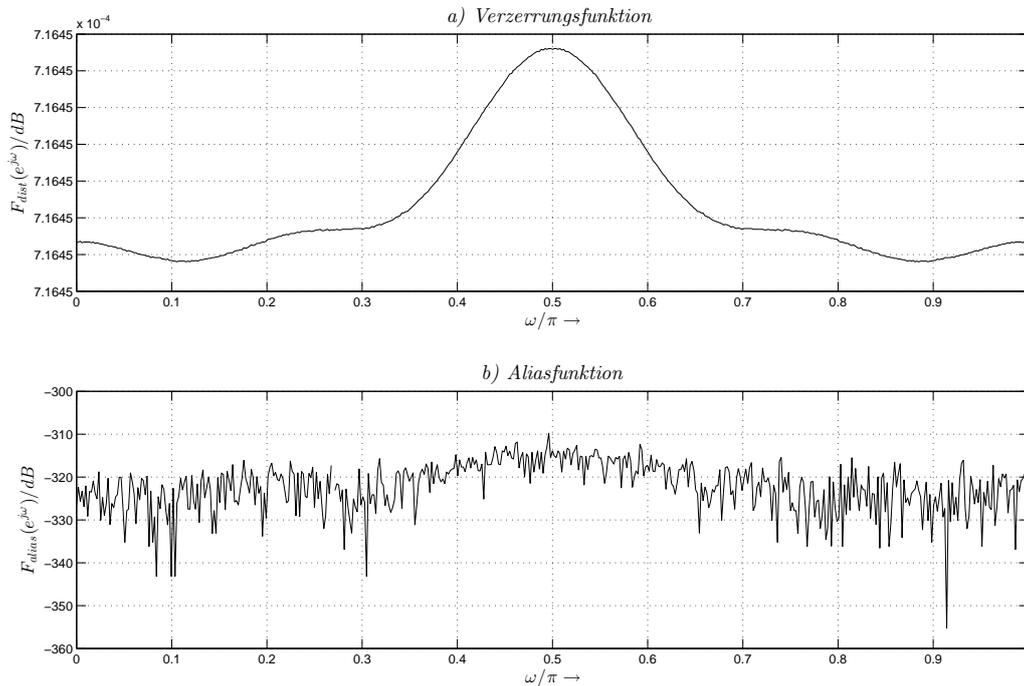


Abbildung 3.6.: Verzerrungs- und Aliasfunktion einer perfekt rekonstruierenden Zweikanal Filterbank

3.4. M-Kanal Filterbänke

Im Folgenden werden die gängigsten Filterbankstrukturen aufgezeigt. Darunter fallen unter anderem die Baum- und die Parallelstruktur. Eine weitere Struktur ist durch die komplex modulierte Filterbank gegeben. Da die in dieser Arbeit verwendete Struktur auf den komplex modulierten Filterbänken basiert, werden diese in gesonderten Abschnitten beschrieben.

3.4.1. Baumstruktur

Eine Filterbank in Baumstruktur ist in Ebenen (*engl. Level*) unterteilt. Jede neue Ebene erhöht die Zerlegung des Frequenzbereichs um doppelt so viele Teilbänder bzw. jedes Teilband wird in zwei neue Teilbänder zerlegt. Dabei werden stets abstraten-dezimierte Zweikanal Filterbänke verwendet. Das Ergebnis ist eine äquidistante Zerlegung des Frequenzbandes zu $2^{\text{Anz. Ebenen}}$ Teilbänder. Abbildung 3.7 zeigt das Blockschaltbild einer Ebene 2 Filterbank in Baumstruktur. Beruhen die Teilfilter $H_0^k(z)$,

$H_1^k(z)$, $G_0^k(z)$ und $G_1^k(z)$ auf einer perfekt rekonstruierenden Zweikanal Filterbank, so besitzt auch das komplette System diese Eigenschaft [Vai92]. Der Nachteil, der mit dieser Struktur einhergeht, ist die hohe Rechenkomplexität [Fli93].

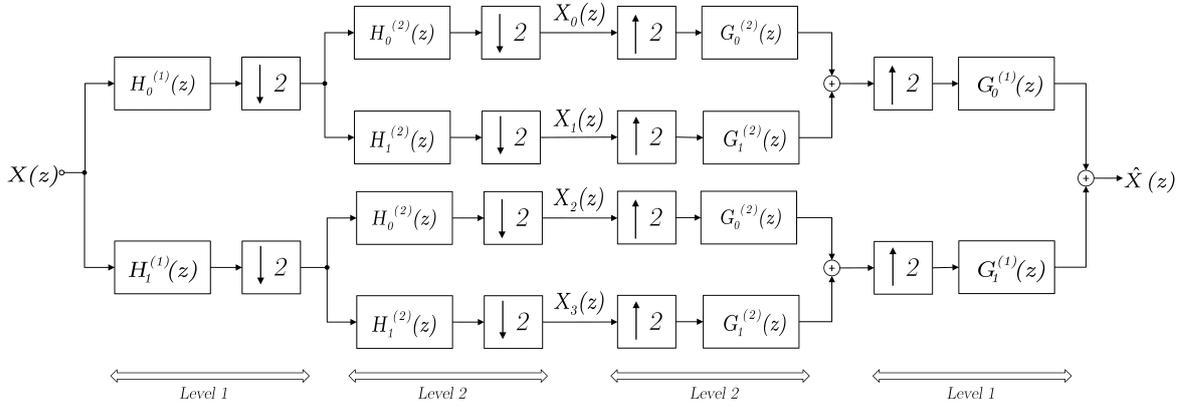


Abbildung 3.7.: Filterbank in Baumstruktur

3.4.2. Parallelstruktur

Im Gegensatz zur Baumstruktur kann mit einer Filterbank in Parallelstruktur eine beliebige Anzahl an Teilbänder erreicht werden. Abbildung 3.8 zeigt das Blockschaltbild einer solchen Filterbank mit M Teilbändern. Da die Anzahl der Teilbänder dem Dezimations- bzw. Interpolationsfaktor entspricht, werden solche Filterbänke auch maximal dezimierte Filterbänke genannt. Auch diese Struktur kann frei von Aliasing sein und sogar perfekte Rekonstruktion erreichen. Die Entwicklung einer solchen Filterbank ist jedoch sehr komplex und der Rechenaufwand steigt überproportional mit steigender Anzahl an Teilbändern [Fli93].

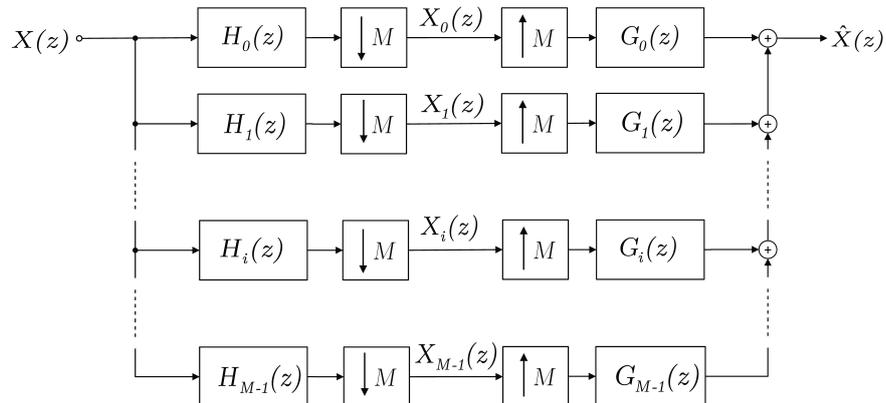


Abbildung 3.8.: Filterbank in Parallelstruktur

3.5. DFT-Filterbank

Durch einen moderaten Rechenaufwand bietet die Kategorie der komplex modulierten Filterbänke basierend auf dem DFT-Algorithmus eine attraktive Alternative [Fli93]. M-Analysefilter entstehen durch äquidistante Frequenzverschiebung eines geeigneten Prototypenfilters $H_0(e^{j\omega})$. Die Teilfilter besitzen dabei die Form

$$H_i(e^{j\omega}) = H_0(e^{j(\omega-2\pi i/M)}) , i = 0, 1, 2, \dots, M - 1. \quad (3.16)$$

Mit der z-Transformation und dem Modulationsoperator $W_M = e^{-j2\pi/M}$ folgt für das i-te Analyse-Filter

$$H_i(z) = H_0(zW_M^i) , i = 0, 1, 2, \dots, M - 1. \quad (3.17)$$

Damit der Frequenzbereich $0 \leq \omega \leq 2\pi$ in M Teilbänder aufgeteilt werden kann, wird für das Prototypenfilter ein 1/M-Bandfilter⁵ gewählt, vgl. Abbildung 3.9 a).

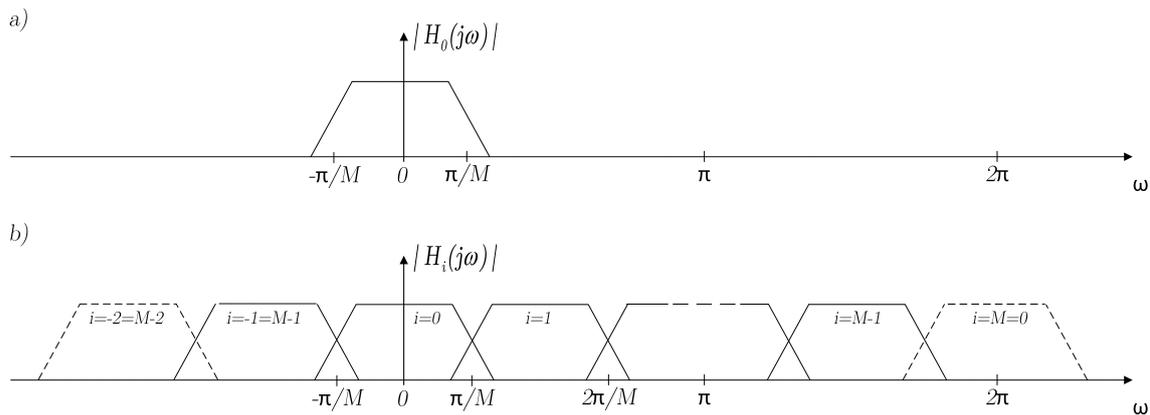


Abbildung 3.9.: Darstellung der äquidistanten Frequenzverschiebung eines Prototypen mittels DFT Algorithmus

Die Verschiebung des Prototypen wird durch den Modulationsfaktor W_M erreicht. Aufgrund der Periodizität um 2π ist beispielsweise das Teilfilter $H_{M-1}(e^{j\omega})$ identisch mit dem Teilfilter $H_{-1}(e^{j\omega})$, vgl. Abbildung 3.9 b). Dabei ist zudem zu erkennen, dass es sich hierbei um keine gleichmäßige Aufteilung des (Nutz-)Frequenzbandes handelt, da sich das Filter $H_0(e^{j\omega})$ in Basisbandlage befindet. Aufgrund des Nyquist-

⁵Ein 1/M-Bandfilter, oder auch Nyquistfilter, ist ein Filter, welches die Durchlass- und Stopbandfrequenz symmetrisch um π/M besitzt. Zudem ist jeder M-te Koeffizient gleich null, ausgenommen des mittleren, dieser hat den Wert 1/M. Für den Fall M=2 handelt es sich bspw. um ein sogenanntes Halbband-Filter [Sau07].

Frequenzbereiches ($0 \leq \omega \leq \pi$) zerlegt eine M -DFT-Filterbank den Nutz-Frequenzbereich in $M/2$ Bänder. Bisher wurde der DFT-Algorithmus

$$y[i] = \sum_{n=0}^{M-1} x[n] \cdot W_M^{in}, \quad 0 \leq i \leq M-1. \quad (3.18)$$

nicht erwähnt bzw. verwendet. Damit eine DFT-Filterbank entsteht, ist die Polyphasenzerlegung entscheidend. Die in Anhang A beschriebene Polyphasenzerlegung kann dazu benutzt werden ein Filter in M Polyphasen zu zerlegen

$$H_0(z) = \sum_{\lambda=0}^{M-1} z^{-\lambda} \cdot E_\lambda(z^M). \quad (3.19)$$

Eine Substitution von $z \rightarrow zW_M^\lambda$ ergibt

$$H_0(z) = \sum_{\lambda=0}^{M-1} z^{-\lambda} W_M^{-\lambda} \cdot E_\lambda(z^M W_M^M). \quad (3.20)$$

Mit $W_M^M = 1$ und $i = 0, 1, \dots, M-1$ folgt für die Analyse-Teilfilter

$$H_i(z) = \sum_{\lambda=0}^{M-1} z^{-\lambda} W_M^{-i\lambda} \cdot E_\lambda(z^M W_M^{iM}). \quad (3.21)$$

Der Zusammenhang zwischen Gleichung 3.21 und Gleichung 3.18 besteht darin, dass $x[n]$ den Polyphasen E_λ und $y[i]$ den Teilfiltern H_i entspricht. Zudem ist der Modulationsoperator komplex-konjugiert. Dies beschreibt, mit Ausnahme des Vorfaktors $1/M$, den Algorithmus der inversen DFT. Abbildung 3.10 zeigt das Blockschaltbild einer DFT-Filterbank für eine beliebige Anzahl M an Teilbändern.

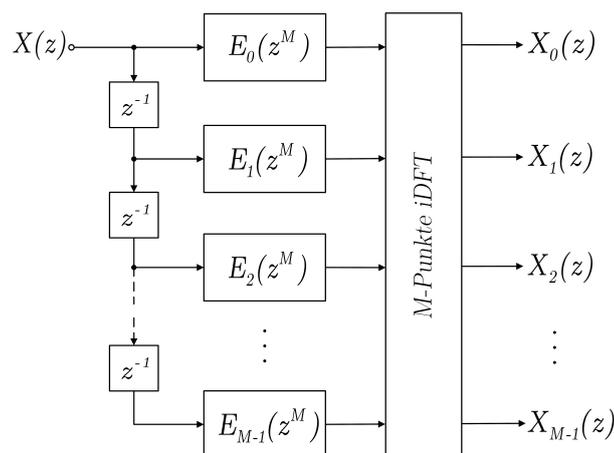


Abbildung 3.10.: Blockschaltbild einer DFT-Analyse-Filterbank

Die Ausgangssignale der Analyse-Filterbank ergeben sich zu

$$\begin{bmatrix} X_0(z) \\ X_1(z) \\ X_2(z) \\ \vdots \\ X_{(M-1)}(z) \end{bmatrix} = \mathbf{D}^* \cdot \begin{bmatrix} E_0(z^M) \\ z^{-1}E_1(z^M) \\ z^{-2}E_2(z^M) \\ \vdots \\ z^{-(M-1)}E_{(M-1)}(z^M) \end{bmatrix} \cdot X(z), \quad (3.22)$$

dabei ist \mathbf{D}^* die komplex-konjugierte DFT-Matrix

$$\mathbf{D} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & W_M^1 & W_M^2 & \dots & W_M^{(M-1)} \\ 1 & W_M^2 & W_M^4 & \dots & W_M^{2(M-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_M^{(M-1)} & W_M^{2(M-1)} & \dots & W_M^{(M-1)(M-1)} \end{bmatrix} \quad (3.23)$$

mit dem Modulationsoperator $W_M = e^{-j\frac{2\pi}{M}}$. Die Synthese-Filterbank, bestehend aus DFT-Matrix und Polyphasenkomponenten des Prototypen bilden die Signalsynthese

$$\hat{X}(z) = \begin{bmatrix} X_0(z) \\ X_1(z) \\ X_2(z) \\ \vdots \\ X_{(M-1)}(z) \end{bmatrix} \cdot \left(\mathbf{D} \cdot \begin{bmatrix} E_0(z^M) \\ z^{-1}E_1(z^M) \\ z^{-2}E_2(z^M) \\ \vdots \\ z^{-(M-1)}E_{(M-1)}(z^M) \end{bmatrix} \right)^T. \quad (3.24)$$

Abbildung 3.11 zeigt das Blockschaltbild der Synthese-Filterbank.

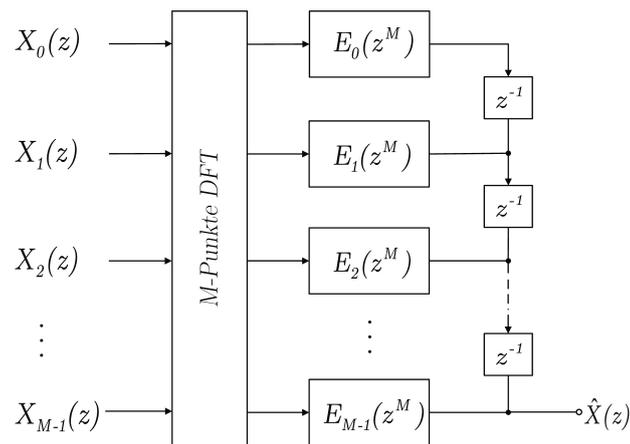


Abbildung 3.11.: Blockschaltbild einer DFT-Synthese-Filterbank

Folgend wird die DFT-Filterbank anhand eines Entwurfs für $M = 4$ untersucht. Das FIR-Prototypenfilter wird mit der Equiripple Methode von Parks-McClellan⁶ entworfen. Die Durchlassgrenzfrequenz wird symmetrisch um π/M zur Sperrfrequenz gewählt ($1/M$ -Band Filter). Die Dämpfungen werden mit 60 dB angenommen. Der Algorithmus liefert eine Ordnung von 66, vgl. Abbildung 3.12. Mit Ausnahme des mittleren Koeffizienten besitzt die Impulsantwort im Abstand von $M = 4$ Nullstellen, was typisch ist für $1/M$ -Bandfilter ist [Sau07].

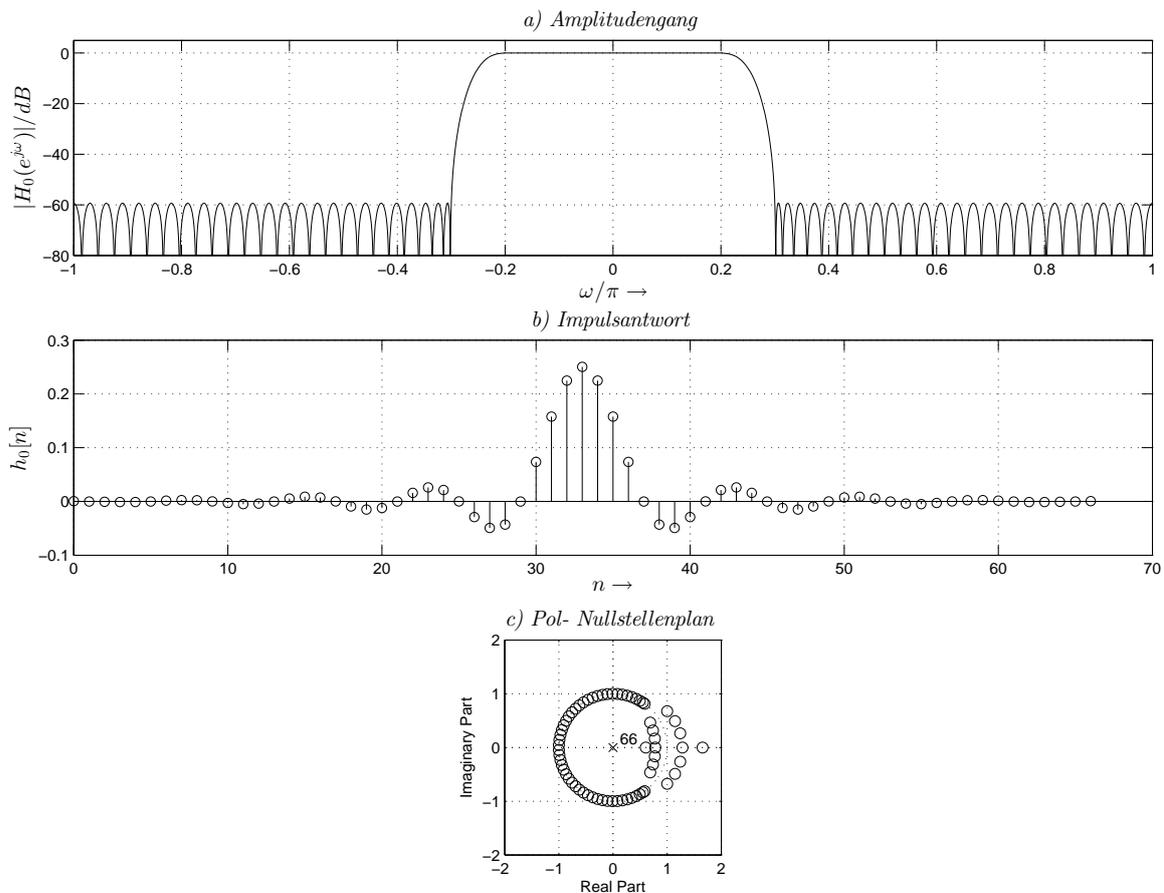


Abbildung 3.12.: Prototypenfilter für eine $M=4$ DFT-Filterbank

Abbildung 3.13 zeigt neben den Amplitudengängen der Teilfilter die Verzerrungs- und Aliasing-Funktion nach Gleichung 3.12 und Gleichung 3.14 der Filterbank. Unschwer

⁶Das Parks-McClellan Design ist eine numerische Methode, welche sich dem Remez Austausch-Algorithmus bedient um die Koeffizienten eines linear-phasigen FIR Filters iterativ zu bestimmen. Dabei wird eine gleichmäßige Welligkeit (Equiripple) im Durchlass- und Sperrbereich erreicht [Fli93].

zu erkennen sind die Einbrüche von 3 dB bei der Verzerrungsfunktion an den Übergangsbereichen der Teilfilter. Zudem tritt starkes Aliasing an diesen Stellen auf. Es handelt sich offensichtlich nicht um eine Filterbank mit PR-Eigenschaft. In praktischen Anwendungen finden DFT-Filterbänke kaum Verwendung, da der Anspruch auf aliasing- und verzerrungsarmen Filterbänken liegt.

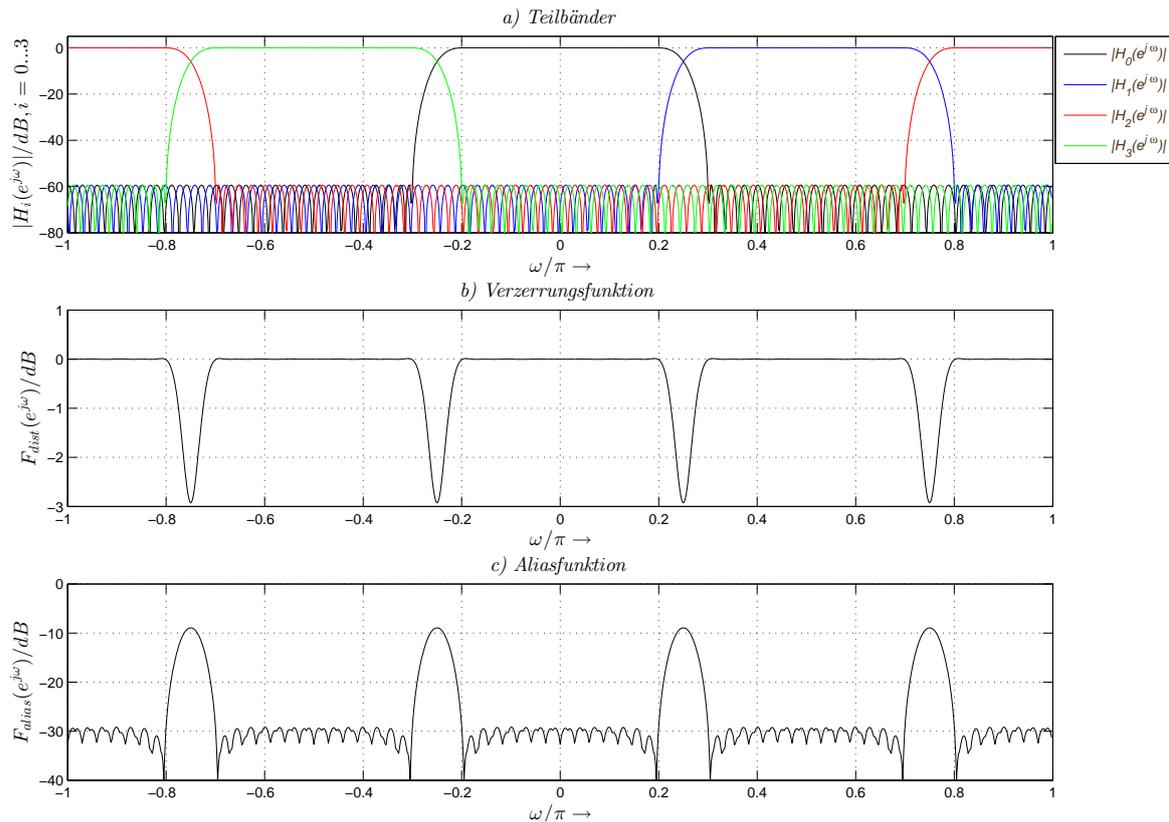


Abbildung 3.13.: Simulationsergebnisse einer M=4 DFT-Filterbank

3.6. Cosinus-modulierte-Filterbank

Die cosinus-modulierte-Filterbank zählt zu den sogenannten Pseudo-Quadrature Mirror Filter (QMF)-Bänken und bietet eine attraktive Alternative zur DFT-Filterbank. Der Begriff Pseudo-QMF entsteht dadurch, dass die Teilfilterkanäle durch äquidistantes Frequenzversetzen eines Prototypen entstehen und dessen Übertragungsfunktionen in guter Näherung leistungskomplementär⁷ sind [Fli93]. Die Herleitung geht über die

⁷Leistungskomplementarität besteht, wenn sich die Summe der quadrierten Betragsfrequenzgänge der Analyse- sowie Synthese-Teilfilter bei allen Frequenzen zu eins ergibt [Fli93].

DFT-Filterbank wodurch sie zur Kategorie der komplex modulierten Filterbank zählt. Eine wesentliche Problematik liegt im ausgeprägtem Aliasing des Übergangsbereichs der Teilbänder, der Einbrüche in der Verzerrungsfunktion sowie die nicht einheitlichen Bandbreiten der Teilfilter. Der Entwurf einer Cosinus-modulierten Filterbank gleicht zunächst dem der DFT-Filterbank, mit dem Unterschied, dass vorerst die Substituierung $M \rightarrow 2M$ stattfindet. Damit folgt für das k -te Teilband

$$P_k(e^{j\omega}) = P_0(e^{j(\omega - \frac{k\pi}{M})}), \quad 0 \leq k \leq 2M - 1 \quad (3.25)$$

bzw. als z-transformierte

$$P_k(z) = P_0(W_{2M}^k z). \quad (3.26)$$

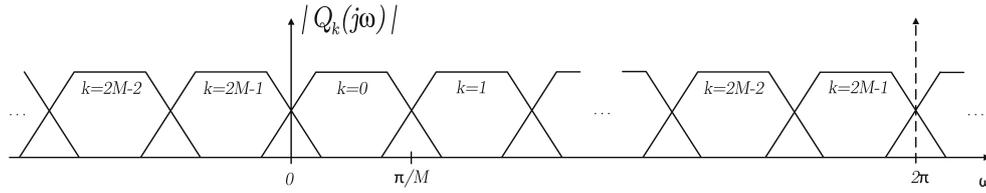
Dies verdoppelt zunächst nur die Anzahl der Teilbänder und verringert den Übergangsbereich des Prototypen auf $\pi/(2M)$. Damit der zu analysierende Frequenzbereich gleichmäßig aufgeteilt wird, bedarf es einer Verschiebung nach rechts bzw. einer Modulation des Filterbank-Amplitudengangs um $\pi/(2M)$. Dies wird durch die Substituierung $z \rightarrow zW_{2M}^{1/2}$ realisiert. Dadurch ändert sich auch das Argument der Polyphasenteilfilter von $E_k(z^{2M})$ mit $(zW_{2M}^{1/2})^{2M} = z^{2M}e^{-j2\pi M/2M}$ zu $E_k(-z^{2M})$. Die Teilfilterausgänge können als Modulation des Prototypen $P_0(z)$ durch $W_{2M}^{1/2}$ geschrieben werden zu

$$Q_k(z) = P_0(zW_{2M}^{k+1/2}) \quad (3.27)$$

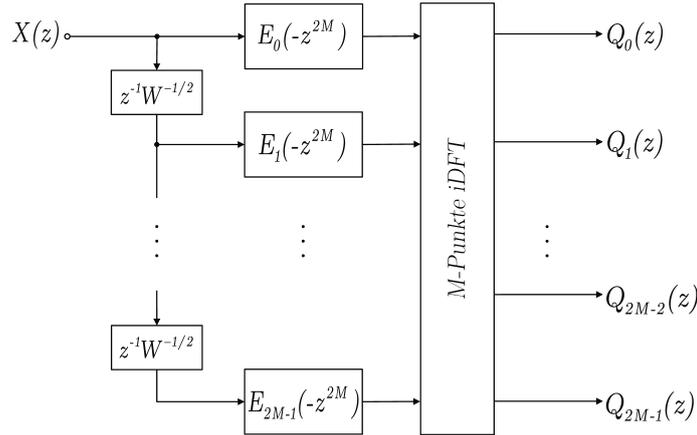
vgl. Abbildung 3.14 b). Durch die Rechts-Verschiebung des gesamten Filterbank-Amplitudengangs sind die Teilamplitudengänge symmetrisch um die Frequenz Null $|Q_k(e^{j\omega})| = |Q_{2M-1-k}(e^{-j\omega})|$, vgl. Amplitudengang-Darstellung in Abbildung 3.14 a). Die Koeffizienten von $Q_k(z)$ verhalten sich demnach komplex-konjugiert zueinander

$$Q_{2M-1-k}(z) = Q_{k,*}(z), \quad 0 \leq k \leq 2M - 1. \quad (3.28)$$

Gleichung 3.28 liefert noch keine gute Lösung, da Aliasing noch nicht eliminiert wird und zudem komplexe Bandpasskoeffizienten resultieren.



(a) Ausgangs-Teilbänder



(b) Blockschaltbild

Abbildung 3.14.: Gleichmäßige Aufteilung des Frequenzbereichs durch Modulation des Prototypen

Für die weitere Herleitung folgen einige Definitionen, die aus [Vai92, S. 357 ff] entnommen sind:

$$U_k(z) = c_k P_0(zW_{2M}^{k+1/2}) = c_k Q_k(z) \quad (3.29)$$

$$V_k(z) = c_k^* P_0(zW_{2M}^{-(k+1/2)}) = c_k^* Q_{2M-1-k}(z) = c_k^* Q_{k,*}(z) \quad (3.30)$$

Mit Gleichung 3.29 und Gleichung 3.30 können die Analyseteilfilter geschrieben werden zu

$$H_k(z) = a_k U_k(z) + a_k^* V_k(z), \quad 0 \leq k \leq M-1. \quad (3.31)$$

Die Syntheseteilfilter ergeben sich zu

$$G_k(z) = b_k U_k(z) + b_k^* V_k(z), \quad 0 \leq k \leq M-1. \quad (3.32)$$

Es kann gezeigt werden, dass bei FIR-Filtern mit geschickter Wahl der Koeffizienten c_k , a_k und b_k eine Auslöschung des Aliasing erreicht werden kann. Eine perfekte Rekonstruktion kann jedoch nicht vollständig realisiert werden, da Amplituden-Verzerrungen werden zunächst nicht beachtet. Auf eine genauere Herleitung der Koeffizienten wird

an dieser Stelle verzichtet und stattdessen auf [Vai92] verwiesen. Die Wahl der Koeffizienten wird im Folgenden aufgestellt. Zunächst gilt

$$a_k = e^{j\Theta_k}, \text{ mit } \Theta_k = (-1)^k \cdot \frac{\pi}{4}. \quad (3.33)$$

Mit der Beziehung

$$a_k = b_k^* \quad (3.34)$$

für den Entwurf der Synthese-Filterbank wird gewährleistet, dass Aliasing unter Verwendung von FIR-Filter mit linearer Phase eliminiert wird. Weiterhin gilt

$$c_k = W_{2M}^{(k+1/2) \cdot N/2}, \quad (3.35)$$

wobei N die Ordnung des Prototypen darstellt. Der erste Term aus Gleichung 3.31 kann nun geschrieben werden zu

$$a_k \cdot U_k(z) = e^{j\Theta_k} c_k P_0(zW_{2M}^{k+1/2}) = d_k Q_k(z) \quad (3.36)$$

und der zweite Term zu

$$a_k^* \cdot V_k(z) = e^{-j\Theta_k} c_k^* P_0(zW_{2M}^{k+1/2}) = d_k^* Q_{2M-1-k}(z). \quad (3.37)$$

Für die Analyse-Filterbank folgt schließlich

$$H_k(z) = d_k Q_k(z) + d_k^* Q_{2M-1-k}(z), \quad (3.38)$$

welches das Blockschaltbild in Abbildung 3.15 verdeutlicht.

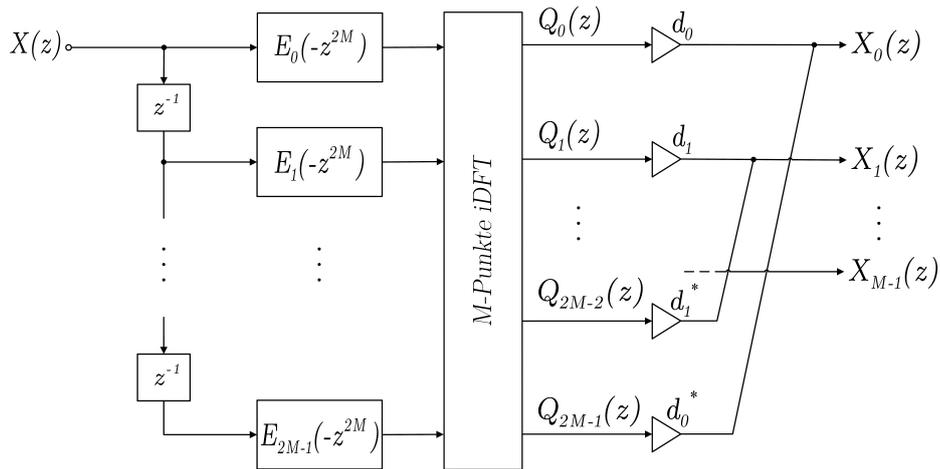


Abbildung 3.15.: Blockschaltbild einer M-Kanal Cosinus-modulierten Analyse-Filterbank

Um die Filterbank explizit entwerfen zu können, bedarf es Kenntnis über die Impulsantwort bzw. der Koeffizienten. Diese können über das Ausschreiben der z -Transformierten von Gleichung 3.36 und Gleichung 3.37 erlangt werden.

$$a_k \cdot U_k(z) = e^{j\Theta_k} W_{2M}^{(k+1/2)N/2} \cdot \sum_{n=0}^N p_0[n] \cdot W^{-(k+1/2)n} z^{-n} \quad (3.39)$$

Die k -te Impulsantwort des ersten Terms ergibt sich zu

$$a_k \cdot u_k[n] = e^{j\Theta_k} W_{2M}^{(k+1/2)N/2} \cdot p_0[n] \cdot W^{-(k+1/2)n}. \quad (3.40)$$

Die Impulsantwort des zweiten Terms aus Gleichung 3.31 erhält man aus der komplexen Konjugation aus obenstehender Gleichung zu

$$a_k^* \cdot v_k[n] = e^{-j\Theta_k} W_{2M}^{-(k+1/2)N/2} \cdot p_0[n] \cdot W^{(k+1/2)n}. \quad (3.41)$$

Somit kann die k -te Teil-Impulsantwort der Analysefilterbank

$$\begin{aligned} h_k[n] &= p_0[n] \cdot \left[e^{-j\Theta_k} e^{-j(k+1/2)\frac{N}{2}\frac{\pi}{M}} e^{j(k+1/2)n\frac{\pi}{M}} + e^{j\Theta_k} e^{j(k+1/2)\frac{N}{2}\frac{\pi}{M}} e^{-j(k+1/2)n\frac{\pi}{M}} \right] \\ h_k[n] &= 2 \cdot p_0[n] \cdot \cos \left[\frac{\pi}{M} \left(k + \frac{1}{2} \right) \left(n - \frac{N}{2} \right) + \Theta_k \right] \end{aligned} \quad (3.42)$$

aufgestellt werden. Aus oben stehender Gleichung resultiert auch der Name dieser Filterbank, da es sich offensichtlich um eine Modulation der Prototypen-Koeffizienten mit einem Cosinusterm handelt. Zudem wird ersichtlich, dass es sich um eine reelle Impulsantwort handelt. Die k -te Teil-Impulsantwort der Synthesefilterbank ergibt sich mit Gleichung 3.34 und Gleichung 3.32 zu

$$g_k[n] = 2 \cdot p_0[n] \cdot \cos \left[\frac{\pi}{M} \left(k + \frac{1}{2} \right) \left(n - \frac{N}{2} \right) - \Theta_k \right]. \quad (3.43)$$

Mit der Zerlegung der Koeffizienten des Prototypen $p_0(n)$ in die einzelnen Polyphasenkomponenten $E_k(-z^{2M})$ und Gleichung 3.42 folgt für die z -Transformierte der Analyseteilfilter

$$H_k(z) = \sum_{n=0}^{2M-1} t_{kn} z^{-n} E_n(-z^{2M}) \quad (3.44)$$

wobei

$$t_{kn} = 2 \cdot \cos \left[\frac{\pi}{M} \left(k + \frac{1}{2} \right) \left(n - \frac{N}{2} \right) - \Theta_k \right] \quad (3.45)$$

$$(3.46)$$

die Elemente der $2M \times M$ Modulationsmatrix \mathbf{T} darstellen, vgl. Abbildung 3.16.

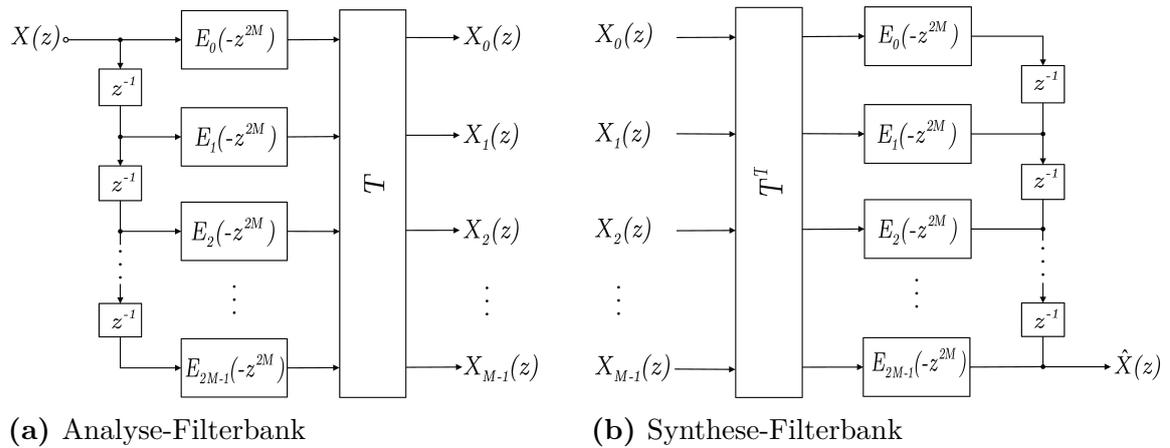


Abbildung 3.16.: Finales Blockschaltbild der Cosinus-modulierten Filterbank

Die Dimensionierung ist somit nur noch vom Prototypen $P_0(z)$ bzw. dessen Polyhasenkomponenten $E_k(-z^{2M})$ abhängig.

Als Beispiel wird eine $M = 4$ Cosinus-modulierte Filterbank entworfen. Um das Ergebnis mit der DFT-Filterbank vergleichen zu können, wird der Prototyp identisch entworfen. Wie bereits beschrieben wird hierbei lediglich $M \rightarrow 2M$ substituiert, sodass ein $1/2M$ -Band Filter resultiert. Abbildung 3.17 zeigt das mit dem Parks-McClellan-Algorithmus entworfene Prototypenfilter. Mit Ausnahme des mittleren Koeffizienten sind der Impulsantwort nun Nullstellen im Raster $2M$ zu entnehmen. Die Ordnung beträgt wie bei dem Prototypen der DFT-Filterbank 66.

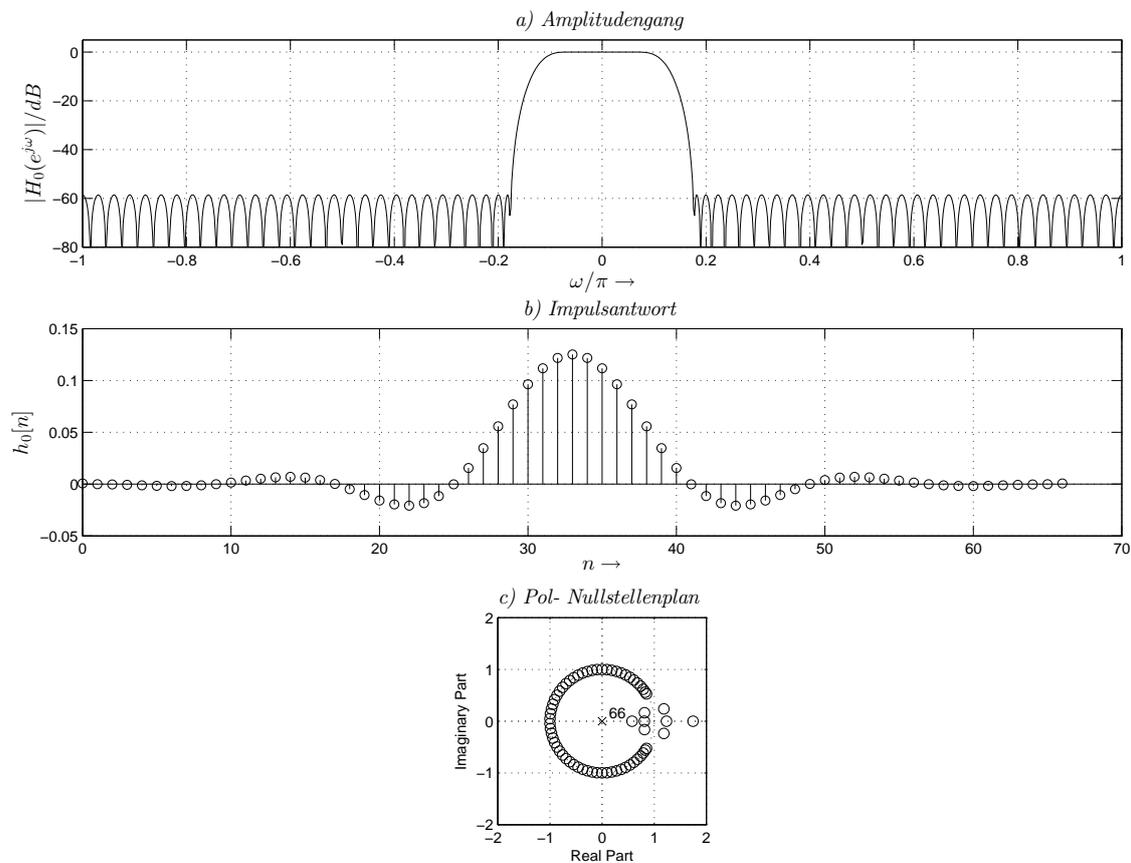


Abbildung 3.17.: Prototypenfilter für eine $M=4$ Cosinus-modulierte Filterbank

Die Amplitudengänge der Teilfilter sowie die Verzerrungs- und Aliasfunktion sind in Abbildung 3.18 a), b) und c) dargestellt. Der Frequenzbereich ($0 \leq \omega \leq \pi$) wird nun gleichmäßig in $M=4$ Teilbänder aufgeteilt. Die Verzerrungsfunktion der Filterbank liefert noch keine wahrnehmbare Verbesserung im Vergleich zur DFT-Filterbank. Die Aliasfunktion zeigt hingegen auch eine in den Übergangsbereichen stärkere Dämpfung. Dies wird dadurch erreicht, dass die Bandbreite halbiert wurde (W_{2M}) und die Modulationsschrittweite der Frequenzverschiebung durch die iDFT unverändert bleibt (W_M). So tritt keine Überlappung der direkt-modulierten Spektren mehr auf, wie es bei der DFT-Filterbank der Fall war. Abbildung 3.19 a) verdeutlicht dies anhand des dritten Ausgangssignals des in der Filterbank enthaltenen iDFT-Blocks. Mit einer Dämpfung von ca. 32 dB liegt die Aliasfunktion, im Vergleich zur Stopbanddämpfung des Prototypen, jedoch noch weit darüber. Dies liegt daran, dass dennoch Aliasing zwischen den Teilbandspektren und dessen modulierten Spiegelspektren und umgekehrt auftritt. Die Darstellung in Abbildung 3.19 b) und c) sollen dies anhand des dritten $H_2(e^{j\omega}) = d_2 \cdot Q_2(e^{j\omega}) + d_2^* Q_{2M-1-2}(e^{j\omega})$ und vierten

$H_3(e^{j\omega}) = d_3 \cdot Q_3(e^{j\omega}) + d_3^* Q_{2M-1-3}(e^{j\omega})$ (vgl. Gleichung 3.38) Analyseteilfilters veranschaulichen. Da es sich im Gegensatz zur DFT-Filterbank um reelle Impulsantworten der Teilbänder handelt, ist das Spektrum spiegelsymmetrisch. Jedes Teilband tritt symmetrisch um die Nullachse bei Frequenzen $\pm\omega$ auf. Die Modulation jedes Spektrums führt zu einer Überlappung im Spiegelspektrum und umgekehrt (schraffierter Bereich in Abbildung 3.19 b) und c)). Es bedarf somit einer zusätzlichen Optimierung des Prototypen.

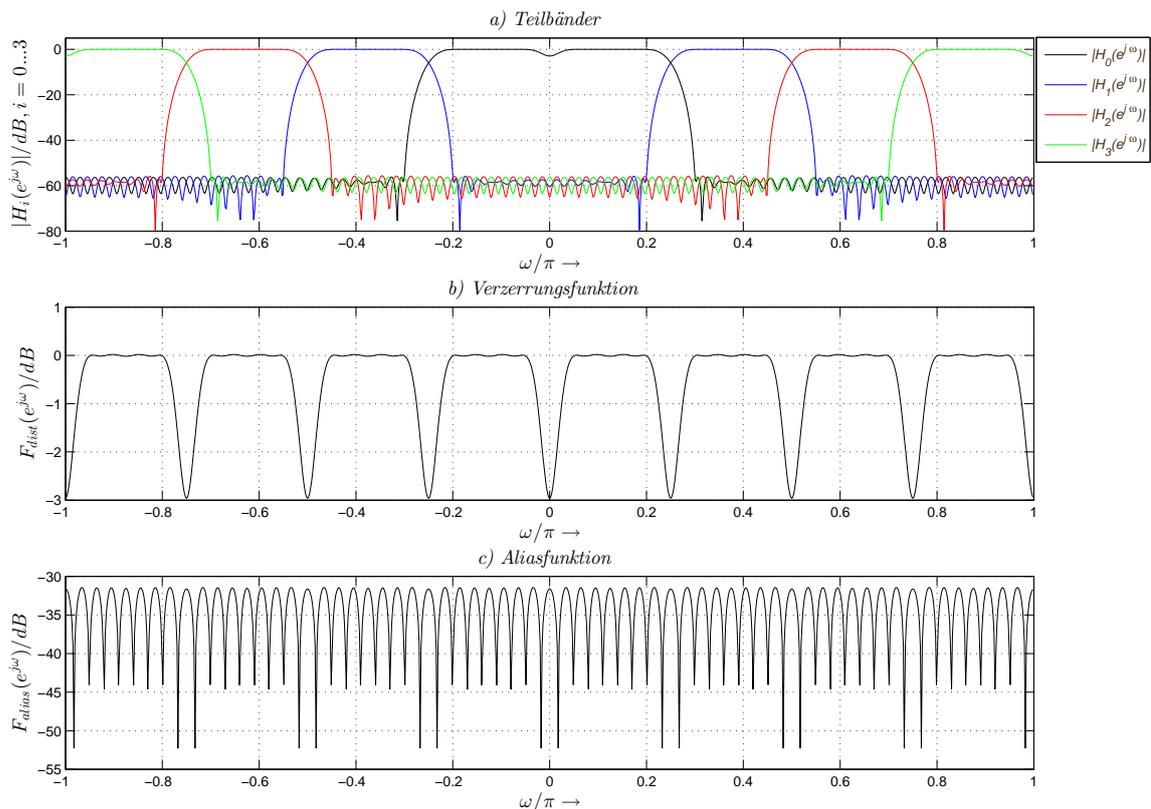
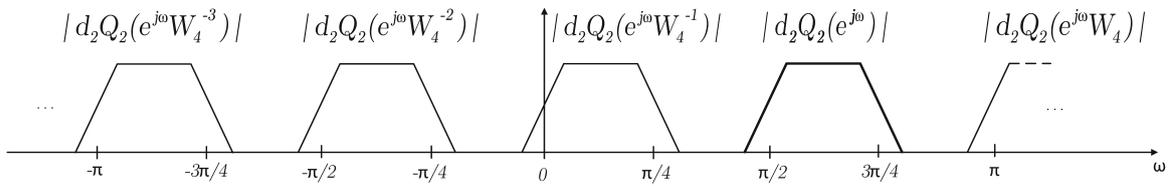
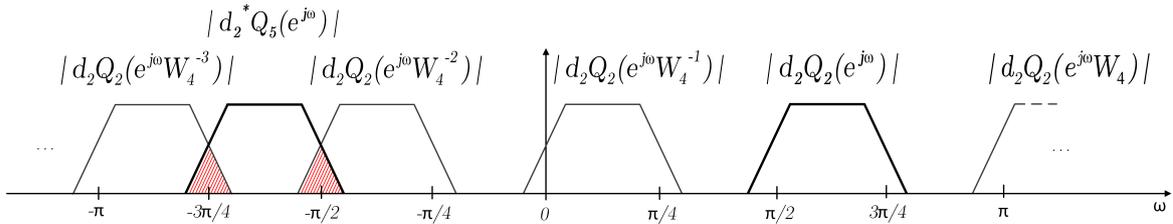


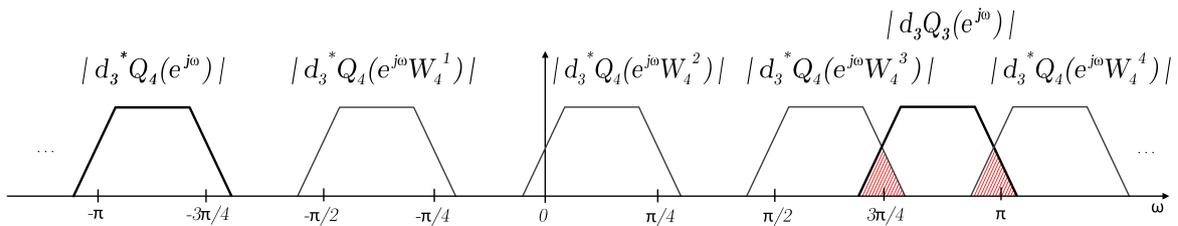
Abbildung 3.18.: Simulationsergebnisse einer $M=4$ Cosinus-modulierten Filterbank



(a) Drittes Teilband ohne komplex-konjugierte Anteile



(b) Drittes Teilband mit einem komplex-konjugierten Anteil – Aliasing entsteht in den rot-schraffierten Bereichen



(c) Viertes Teilband inklusive einem komplex-konjugierten Anteil – Aliasing entsteht in den rot-schraffierten Bereichen

Abbildung 3.19.: Darstellung ausgewählter Teilbänder der Cosinus-modulierten Filterbank für $M=4$, mit zugehörigen Aliaskomponenten äquivalent zum Blockschaltbild Abbildung 3.15

3.6.1. Optimierung

Für eine konstante Verzerrungsfunktion muss das Prototypenfilter im Bereich $0 \leq \omega \leq \pi/M$

$$\left|P_0(e^{j\omega})\right|^2 + \left|P_0(e^{j(\omega-\pi/M)})\right|^2 = 1, \quad (3.47)$$

also leistungskomplementär sein und für $\omega > \pi/M$

$$\left|P_0(e^{j\omega})\right| = 0 \quad (3.48)$$

erfüllen [Mit02]. Ein Entwurf eines Prototypen, welcher die Bedingungen 3.47 und 3.48 erfüllt, ist jedoch nicht realisierbar. Es ist lediglich eine Approximation an diese Bedingungen möglich, welche durch die Zielfunktion (*objective function*)

$$\phi = \max_{0 < \omega < \pi/M} \left\{ \left|P_0(e^{j\omega})\right|^2 + \left|P_0(e^{j(\omega-\pi/M)})\right|^2 - 1 \right\} \quad (3.49)$$

erreicht werden kann. In [Mit02, S. 732 ff] ist ein Verfahren gegeben, welches die Koeffizienten des Prototypen $P_0(e^{j\omega})$ iterativ optimiert, um die Zielfunktion in Gleichung 3.49 zu minimieren. Auf eine detaillierte Beschreibung des Verfahrens wird an dieser Stelle verzichtet und stattdessen auf dessen Quelle referenziert. Das Verfahren benötigt lediglich die Anzahl der Teilbänder M und die Ordnung N des Prototypen. Initialisiert wird die Sperrfrequenz auf $1/M$ und die Durchlassfrequenz zunächst auf $1/4M$. Um die Zielfunktion iterativ zu approximieren, wird der Prototyp in jedem Iterationsschritt mit dem Parks-McClellan-Algorithmus neu entworfen, wobei der Durchlassbereich fünfmal höher als der Sperrbereich gewichtet wird. Die Durchlassfrequenz wird in jedem Schritt näher in Richtung Null bewegt, was die Sperrdämpfung sukzessiv erhöht. Dies geschieht solange, die Abweichung aus Zielfunktion vom aktuellen- und letzten Iterationsschritt größer einem Abbruchkriterium von 10^{-6} ist. Abbildung 3.20 zeigt den optimierten Prototypen. Das Resultat liefert eine maximal-erreichbare Stopbanddämpfung. Diese hat sich bei gleichbleibender Ordnung sogar fast verdoppelt. Dabei wurde der Übergangsbereich breiter und die Sperrfrequenz befindet sich auf $\omega_S = 1/M = 1/4$.

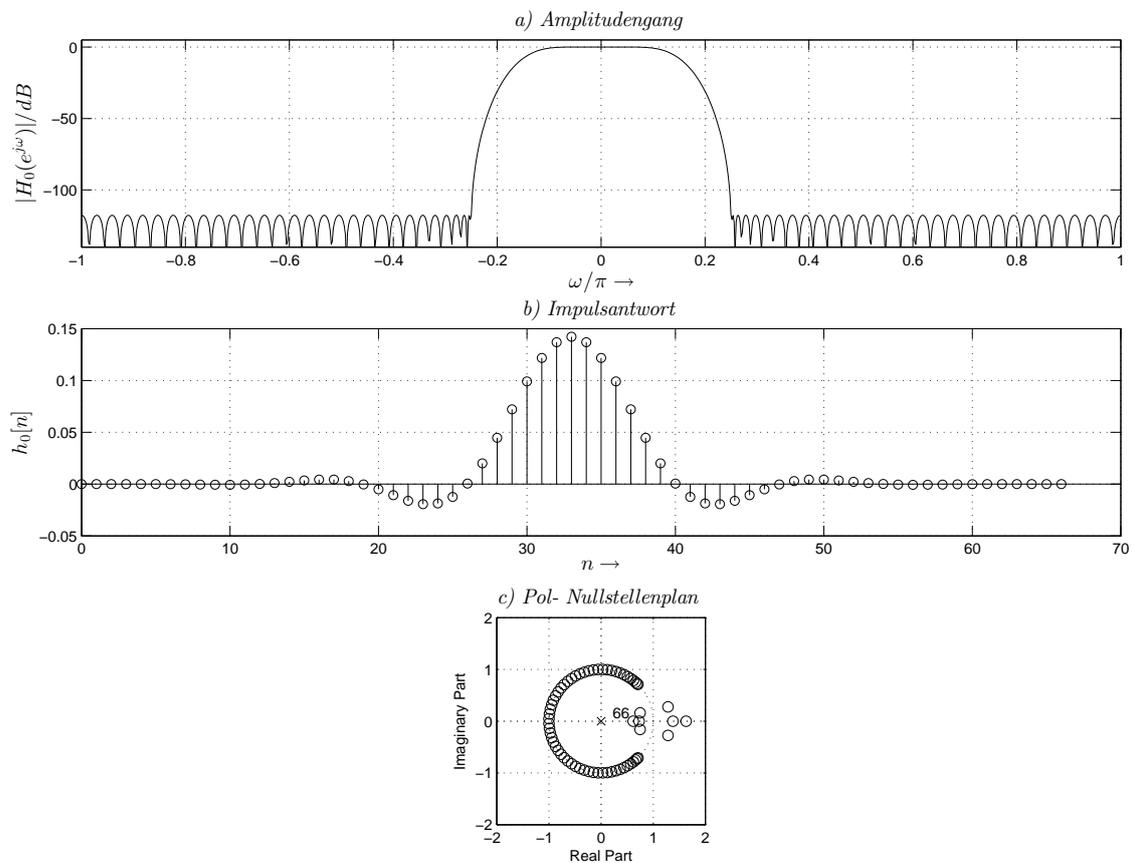


Abbildung 3.20.: Optimiertes Prototypenfilter für eine $M=4$ Cosinus-modulierte Filterbank

Der optimierte Prototyp wird nun für die Cosinus-modulierte Filterbank verwendet. Die Teilbänder sowie die Verzerrungs- und Aliasfunktion sind in Abbildung 3.21 abgebildet. Die Teilfilter besitzen ebenso die Stopbanddämpfung von nahezu 120 dB. Die Verzerrungsfunktion ist nicht konstant, kann jedoch mit einer Abweichung von maximal 0,02 dB für Audioanwendungen vernachlässigt werden. Die Aliasfunktion konnte durch die Optimierung um ca. 25 dB auf insgesamt fast 60 dB an Dämpfung zunehmen.

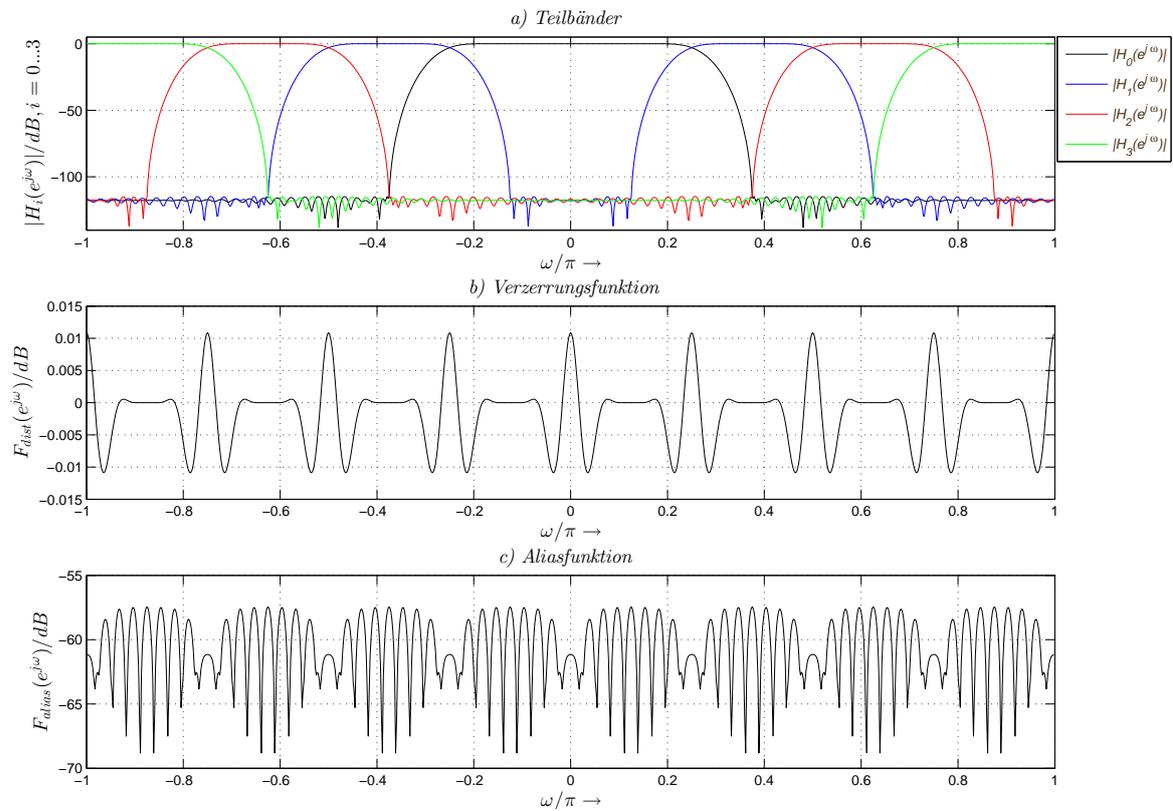


Abbildung 3.21.: Simulationsergebnisse einer $M=4$ Cosinus-modulierten Filterbank mit optimiertem Prototypen

Da der Aufwand der Cosinus-modulierten Filterbank überschaubar ist und sie die perfekte Rekonstruktion so gut annähert, findet sie häufig Anwendung im Bereich der digitalen Audiosignalverarbeitung, so auch im MPEG-1 Verfahren.

4. MPEG-1 Audio Codierung

MPEG-1 ist ein Standard der Moving Picture Experts Group (MPEG) zur verlustbehafteten Video- und Audiodatenkompression. Dieser Standard wurde erstmalig 1991 vorgestellt und ist in der ISO/IEC 11172 festgelegt. Das Ziel der Codierung ist es, Videodaten effizient und qualitativ verlustarm auf einer CD speichern zu können. Darin enthalten sind folgende Abschnitte [Gro14]:

- **Part 1 Systems:** In diesem Teil sind System Definitionen enthalten. So kann es sich beim Decodierprozess um ein gespeichertes Medium oder um einen zu decodierenden Bitstrom (*engl. Stream*) handeln. Des Weiteren wird auf Kompatibilität geprüft, aber auch Daten/Informationen zur Abspielzeit und Bezeichnungen können hier enthalten sein.
- **Part 2 Video:** Der zweite Part beschreibt die Codierung von bewegten Bildern. Diese ähneln sich häufig, sodass die Information nur einmal gespeichert oder übertragen werden muss. Dabei wird bei einer Gruppe von Bildern (GOP: *engl. Group of Pictures*) eine sogenannte Bewegungsprädiktion durchgeführt. Eine zusätzliche Diskrete Cosinus Transformation (DCT) dient zur Dekorrelation der Pixel in einem Bild. Nach einer Quantisierung der Pixel folgt eine VLC-Codierung (*Variable Length Coding*), die zu weiterer Datenreduktion führt.
- **Part 3 Audio:** Die Komprimierung von Audiosignalen beschreibt Part 3. Diese Abschlussarbeit befasst sich ausschließlich mit diesem Teil des MPEG-1 Standards und wird weiter unten fortgeführt.
- **Part 4 Conformance Testing:** Dieser Teil des Standards kann für Entwickler eines Encoders oder Decoders genutzt werden, um zu verifizieren, ob ein gültiger Bitstrom erzeugt wird. Dabei wird ein bestimmter Referenz-Bitstrom verwendet.
- **Part 5 Software Simulation:** Part 5 liefert eine Simulations-Software in C eines Encoders und Decoders, für Part 1 (Systems), Part 2 (Video) und Part 3 (Audio)

Im Folgenden wird die MPEG-1 Audiocodierung (Part 3) erläutert, welche in drei weitere Schichten unterteilt wird:

- **Layer I (MP1):** Pseudo QMF-Bank mit 32 Teilbändern, FFT (512 Punkte), Psychoakustisches Modell 1
- **Layer II (MP2):** Pseudo QMF-Bank mit 32 Teilbändern, FFT (1024 Punkte) Psychoakustisches Modell 1
- **Layer III (MP3):** Pseudo QMF-Bank mit 32 Teilbändern, modifizierte diskrete Cosinus-Transformation jedes Teilbandes, FFT (1024 Punkte), Psychoakustisches Modell 2, Huffman Codierung

Aufgrund der hohen Komplexität des Layer III wird für das zu realisierende System der MPEG-1 Layer I Standard mit einer Abtastfrequenz von $F_s = 48 \text{ kHz}$ verwendet. Alle benutzten Gleichungen sowie Tabellen sind aus [Sta91], [Sal06], [ZÖ8] und [SPA07] entnommen.

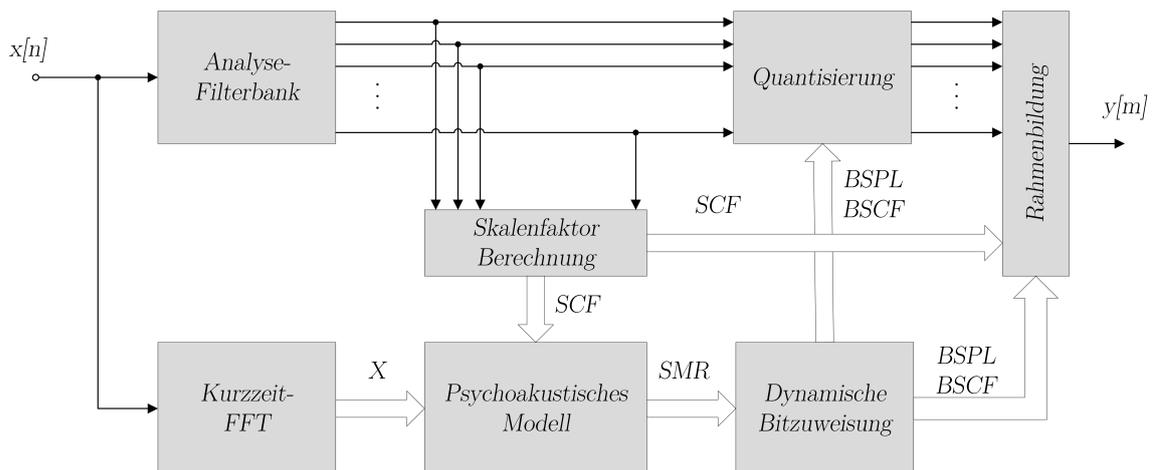


Abbildung 4.1.: Blockschaltbild eines MPEG-1 Layer I und II Encoders

Abbildung 4.1 zeigt das Blockschaltbild eines Encoders. Dieser besitzt sieben wesentliche Funktionsblöcke, wobei das Audiosignal durch $x[n]$ und der komprimierte Bitstrom durch $y[m]$ beschrieben wird. Das MPEG-1 Verfahren stellt mehrere Methoden zur Kopplung des rechten und linken Kanals bereit. Dabei findet beispielsweise die Mid/Side-Stereofonie Anwendung. Diese berechnet aus dem linken und rechten Kanal den Mittenkanal

$$M = \frac{L + R}{2} \quad (4.1)$$

und den Seitenkanal

$$S = \frac{L - R}{2}. \quad (4.2)$$

Die Rekonstruktion des linken Kanals ergibt sich aus der Summe von M und S . Die Rekonstruktion des rechten Kanals aus der Differenz von M und S . In dieser Arbeit wird nur ein Kanal (Monosignal) übertragen. Da kein Verlust des rechten oder linken Kanals auftreten soll, wird das zu komprimierende Audiosignal $x[n]$ aus dem Mittenkanal nach Gleichung 4.1 gebildet.

4.1. Filterbank

Bei MPEG-1 handelt es sich um ein Teilbandcodierungsverfahren (*engl. Subband Coding*), was nichts anderes bedeutet, als dass das Audiosignal in Teilbänder zerlegt wird, welche im Anschluss quantisiert bzw. codiert werden. Dass die Teilbänder gehörgerecht in die Frequenzgruppen aus Tabelle 2.1 zerlegt werden, liegt zwar nahe, zieht jedoch einen sehr hohen Rechenaufwand mit sich. Deshalb wird der Frequenzbereich in äquidistantem Abstand zerlegt. Dies realisiert eine Cosinus-modulierte Pseudo QMF-Bank wie in Abschnitt 3.6 beschrieben. Dabei wird das Audiosignal in $M=32$ Teilbänder mit je der Bandbreite von

$$BW = \frac{F_s/2}{M} = 750 \text{ Hz} \quad (4.3)$$

zerlegt. Unglücklicherweise führt dies dazu, dass jedes der unteren Frequenzbänder der Filterbank mehrere Bänder im Bark-Bereich überlappen bzw. einschließen. Dadurch ist zunächst keine gehörgerechte Teilband-Quantisierung gegeben. Ein gutes Ergebnis wird jedoch erzielt, indem die niedrigste Mithörschwelle aus einer Reihe an Gehör-Frequenzbändern, die durch ein Teilfilterband eingeschlossen werden, gesucht wird und diese für die Quantisierung verwendet wird. Die Suche nach Mithörschwellen in einem Frequenzband wird mittels Fast Fourier Transformation (FFT) des Eingangssignals und einem psychoakustischen Modells erreicht. Dieser Vorgang wird in folgenden Abschnitten beschrieben. Der MP3-Standard nutzt bspw. eine weitere Zerlegung der 32 Teilbänder durch eine diskrete Cosinus-Transformation, wodurch eine feinere Frequenzauflösung erreicht wird, die der Bark-Skala noch näher kommt.

In [Sta91], Seite 79 – Tabelle 3-C.1 sind die 512 reellwertigen Koeffizienten gegeben, die mit dem Flussdiagramm aus Abbildung 4.2 die MPEG-1 Analyse-Filterbank bilden. Abweichend davon wird die Filterbank, basierend auf der Herleitung aus Abschnitt 3.6, realisiert. Die Prozedur ist identisch mit dem Flussdiagramm der MPEG-1 Variante. Unterschiede können lediglich durch die Koeffizienten entstehen. Die Abweichungen dieser sind in Abbildung 4.3 dargestellt. Dabei sind Sprünge in beiden Impulsantworten zu erkennen. Diese gehen aus der Herleitung der Filterbank hervor, wobei das Argument $E_k(-z^{2M})$, eines jeden Polyphasenteilfilters, mit $M=32$ dazu führt, dass al-

le vierundsechzigsten Koeffizienten das Vorzeichen alterniert. Die Abweichungen sind mit bis zu $6 \cdot 10^{-4}$ zwar präsent, jedoch vernachlässigbar.

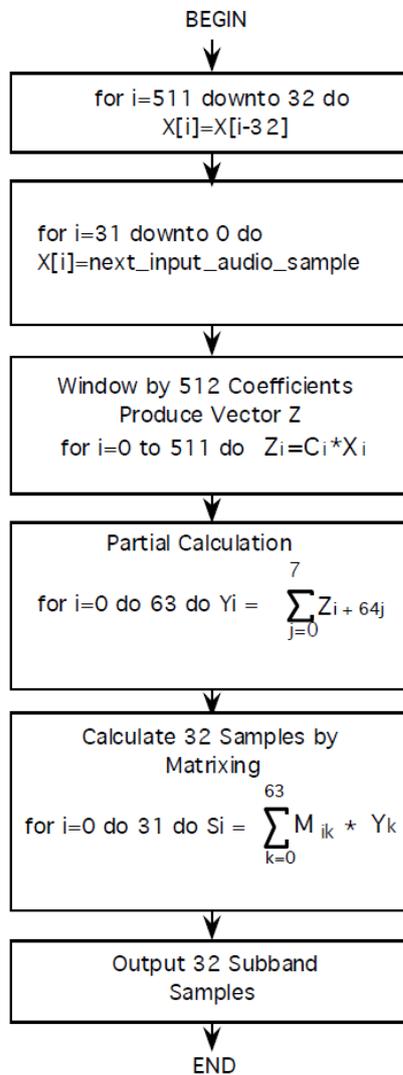


Abbildung 4.2.: Flussdiagramm des Entwurfs der MPEG-1 Analyse-Filterbank [Sta91]

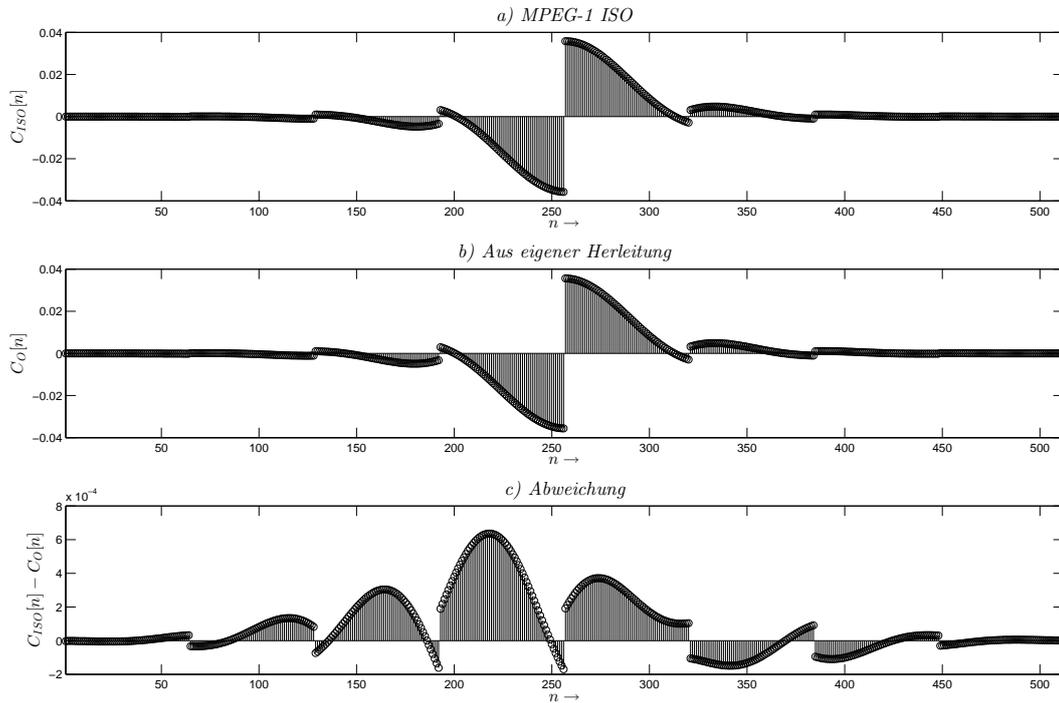


Abbildung 4.3.: Vergleich der in [Sta91] gegebenen Koeffizienten mit denen basierend auf der Methodik aus Abschnitt 3.6 hergeleiteten Koeffizienten

Abbildung 4.4 a) zeigt noch einmal die Koeffizienten des verwendeten Prototypen ohne alternieren des Vorzeichens jeder vierundsechzig Koeffizienten. Abbildung 4.4 b) stellt den Amplitudengang dar. Unter anderem an der Sperrdämpfung bei $\omega_S = \pi/2M = \pi/64$ wird ersichtlich, warum gerade eine Ordnung von 511 gewählt wurde. Mit über 110 dB Dämpfung wird sichergestellt, dass das Audiosignal eindeutig in die 32 einzelnen Teilbänder zerlegt werden kann, ohne dass im Sperrbereich eines jeden Teilbands noch Signalanteile vorhanden sind. Dies ist dadurch begründet, dass das Audiosignal durch einen $l=16$ Bit Analog-Digital-Converter (ADC) linear quantisiert wird, woraus ein maximaler Signal-Rauschabstand (*engl. Signal-to-Noise Ratio (SNR)*) von

$$SNR_{ADC} = 10 \cdot \log_{10} \left(\frac{3}{2} 2^{2l} \right) \approx 1,77 + 6 \cdot l = 97,8 \text{ dB} \quad (4.4)$$

resultiert [Kam11].

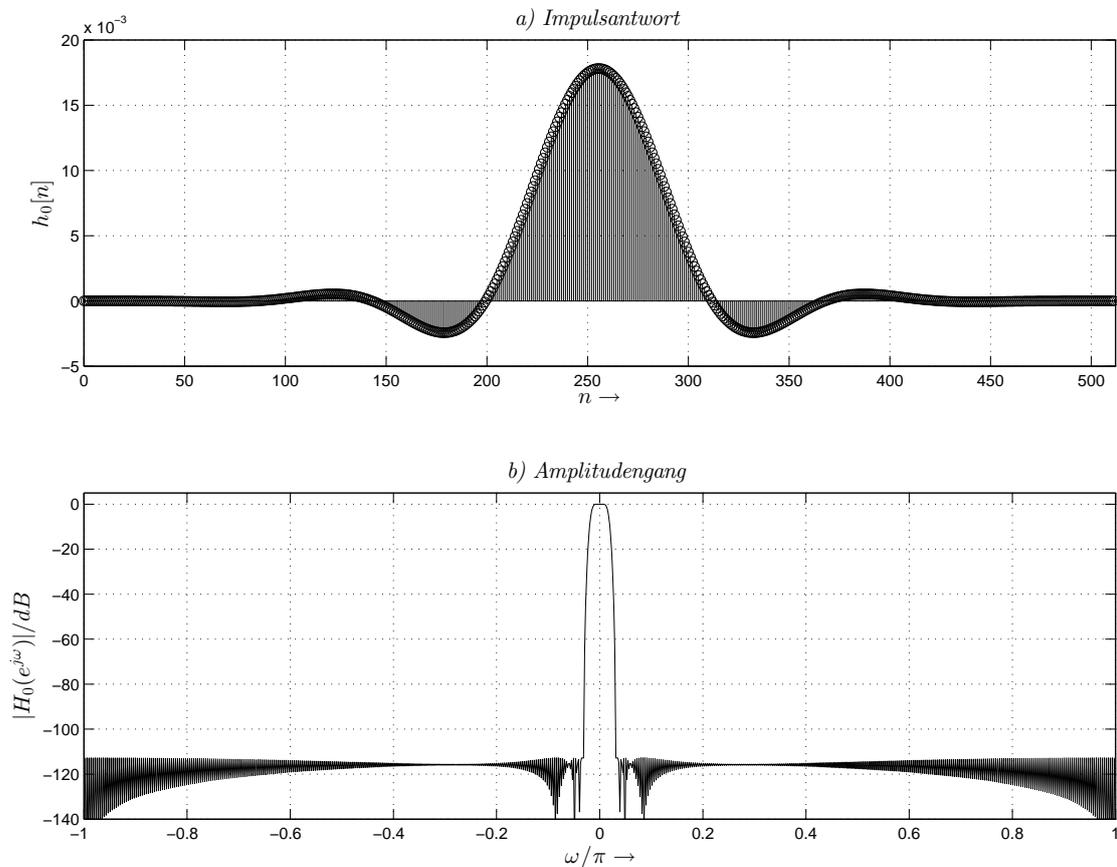


Abbildung 4.4.: Darstellung des verwendeten Filterbank Prototypen

Ein weiterer Grund für eine so hohe Ordnung des Prototypen ist, dass durch die Polyphasendekomposition in $2M = 64$ Polyphasenteilfilter jedes Filter effektiv $512/64 = 8$ Koeffizienten besitzt, woraus eine identische Laufzeit resultiert.

Die Polyphasenzerlegung bringt einen wichtigen Vorteil mit sich. Um den Rechenaufwand klein zu halten, kann eine Abstraten-Reduktion durchgeführt werden, da die Teilfilter jeweils auf z^{2M} arbeiten, vgl. hierzu Abbildung 3.16 für $M=32$. Eine maximal dezimierte Filterbank kann jedoch nicht erreicht werden, da das Prototypenfilter auf $\omega_S = \pi/64$ bandbegrenzt ist und bei einer Dezimation um 64 Alias entsteht. Eine Dezimation um $M=32$ mindert jedoch den Rechenaufwand ebenso erheblich. Dabei kann die Noble-Identity Nr. 1⁸ ausgenutzt werden. Abbildung 4.5 a) zeigt das Blockschaltbild für eine dezimierte Cosinus-modulierte Analyse-Filterbank. Abbildung 4.5 b) zeigt dabei eine Darstellung, die für die spätere Implementierungen anschaulicher

⁸Die Noble Identity Nr. 1 gibt an, dass eine Abstratenreduktion um den Faktor M vor ein Filter in z^M geschaltet werden kann [Sau07].

ist. Der mit der Abtastfrequenz rotierende Schalter bildet die Dezimation und deutet auf die zur Zeit arbeitenden Polyphasenfilter hin.

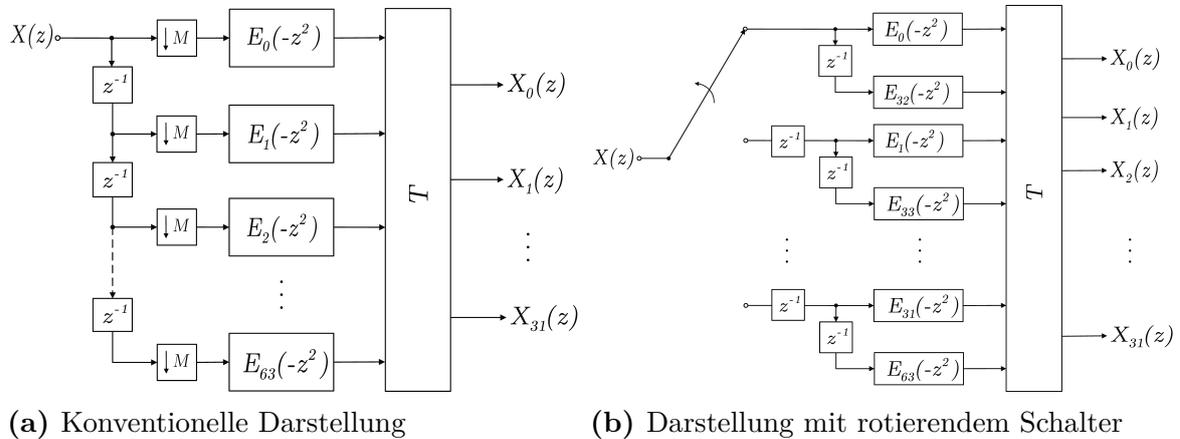


Abbildung 4.5.: Blockschaltbild der Cosinus-modulierten Analyse-Filterbank mit Abwärtstastung

Die Amplitudengänge der Analyse-Filterbank sind in Abbildung 4.6 a) dargestellt. Die Verzerrungsfunktion in Abbildung 4.6 b) kann für Audioanwendungen als konstant angesehen werden. Abbildung 4.6 c) zeigt eine Unterdrückung des Aliasing um bis zu 55,5 dB.

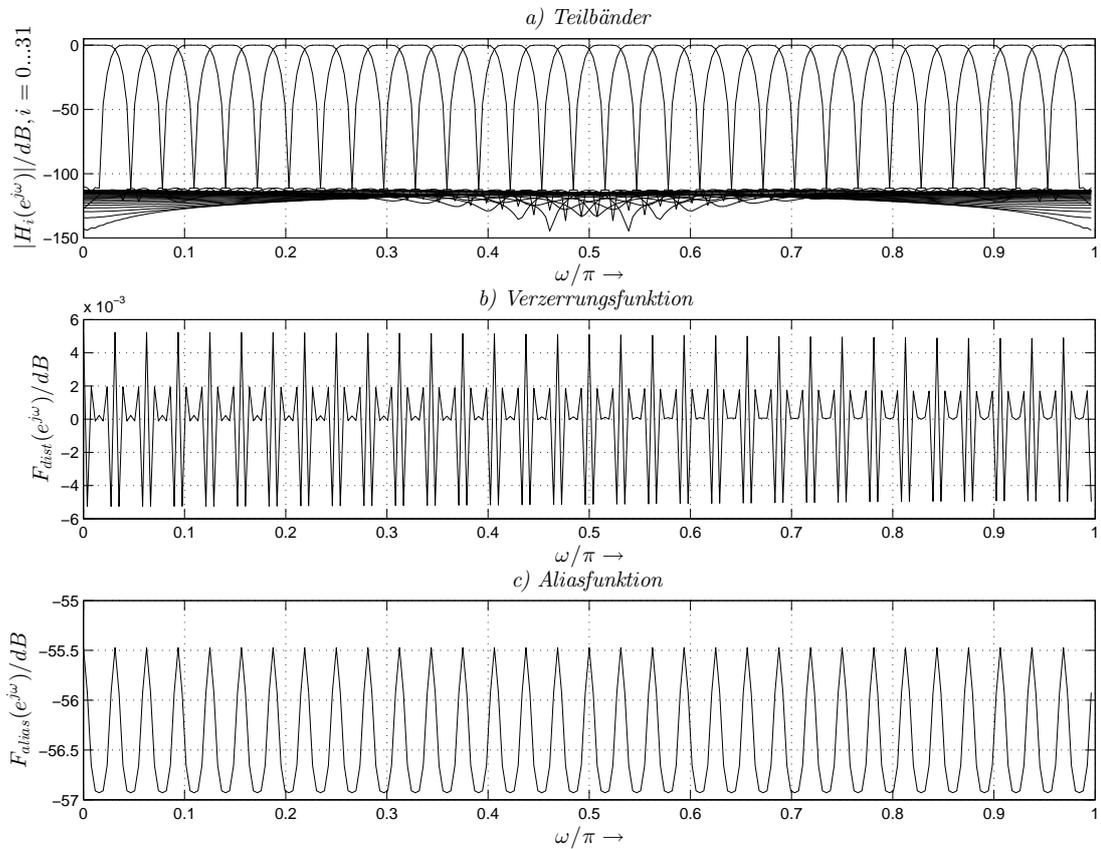


Abbildung 4.6.: Simulationsergebnisse der $M=32$ Cosinus-modulierten Filterbank

4.2. Skalenfaktoren

Die Skalenfaktoren werden anhand von 12 Abtastwerte aus jedem der 32 Teilbänder gebildet. Dafür wird das Betragsmaximum der 12 Abtastwerte gesucht. Der nächstgrößere Wert aus Tabelle 4.1 wird als Skalenfaktor zugewiesen. Dieser dient als Normierung der Teilbandabtastwerte für die Quantisierung. Da der Skalenfaktor immer größer als die 12 Teilbandabtastwerte ist, bleiben die Werte der Teilbänder im Bereich $0 < X(i)/SCF(i) < 1$. Die aufgelisteten Skalenfaktoren können wie folgt berechnet werden:

$$SCF(d) = 2 \cdot \left(\sqrt[3]{\frac{1}{2}} \right)^d, \quad d = 0, 1, 2, \dots, 62 \quad (4.5)$$

Damit der Decoder die Teilbänder korrekt zusammensetzen kann, werden die Skalenfaktoren jedes Teilbands benötigt. Diese müssen daher bei der Rahmenbildung im Header mit aufgeführt werden.

d	SCF	d	SCF	d	SCF
0	2.000000000000000	21	0.015625000000000	42	0.00012207031250
1	1.58740105196820	22	0.01240157071850	43	0.00009688727124
2	1.25992104989487	23	0.00984313320230	44	0.00007689947814
3	1.000000000000000	24	0.007812500000000	45	0.00006103515625
4	0.79370052598410	25	0.00620078535925	46	0.00004844363562
5	0.62996052494744	26	0.00492156660115	47	0.00003844973907
6	0.500000000000000	27	0.003906250000000	48	0.00003051757813
7	0.39685026299205	28	0.00310039267963	49	0.00002422181781
8	0.31498026247372	29	0.00246078330058	50	0.00001922486954
9	0.250000000000000	30	0.001953125000000	51	0.00001525878906
10	0.19842513149602	31	0.00155019633981	52	0.00001211090890
11	0.15749013123686	32	0.00123039165029	53	0.00000961243477
12	0.125000000000000	33	0.000976562500000	54	0.00000762939453
13	0.09921256574801	34	0.00077509816991	55	0.00000605545445
14	0.07874506561843	35	0.00061519582514	56	0.00000480621738
15	0.062500000000000	36	0.000488281250000	57	0.00000381469727
16	0.04960628287401	37	0.00038754908495	58	0.00000302772723
17	0.03937253280921	38	0.00030759791257	59	0.00000240310869
18	0.031250000000000	39	0.00024414062500	60	0.00000190734863
19	0.02480314143700	40	0.00019377454248	61	0.00000151386361
20	0.01968626640461	41	0.00015379895629	62	0.00000120155435

Tabelle 4.1.: Skalenfaktoren

4.3. FFT

Um die Kenntnisse von Zwicker für gehörgerechte Reduzierung der Bitrate anwenden zu können, bedarf es der Information des Leistungspegels. Dieser wird anhand des Spektrums ermittelt. An dieser Stelle muss auf die Eigenschaft eines Audiosignals eingegangen werden. Die diskrete Datenfolge $x[n]$ geht aus dem analogen Audiosignal hervor. Dieses ist aus Sicht der Informationstheorie ein Zufallssignal. Somit ist die zu analysierende Datenfolge ebenso ein Zufallssignal bzw. ein stochastisches Signal. Von diesem wird vorausgesetzt, dass es statistisch unabhängig (unkorreliert) und gleichverteilt sei. Um eine Aussage über die spektrale Leistungsverteilung des Signals treffen zu können, muss das Leistungsdichtespektrum (LDS) berechnet werden. Dieses erhält man allgemein durch die Wiener-Khintchine Beziehung, welche besagt, dass die spektrale Leistungsdichte eines stationären Zufallsprozesses die Fourier-Transformierte der korrespondierenden Autokorrelationsfunktion (AKF) ist [Kam11]. Die AKF führt dazu, dass Signalanteile, die weniger eindeutig sind, kleiner gewichtet werden. Zudem ist sie immer eine gerade Funktion, was bei der Fourier-Transformation, welche

von einem periodischem Signal ausgeht, verfälschende Signalsprünge vermeidet. Dies reduziert den „Leck-Effekt“⁹. Anstatt die AKF zu berechnen ist in der Praxis die Fenstermethode ein gängiges Verfahren [WS04]. Dazu wird die Datenfolge mit einer Fenster-Funktion

$$x_w[n] = x[n] \cdot w[n] \quad (4.6)$$

gewichtet, wodurch eine Ein- und Ausblendung der Datenfolge entsteht. Beim MPEG-1 Standard kommt das Hann-Fenster

$$w_{hann}[n] = 0,5 \cdot \left(1 - \cos\left(2\pi \cdot \frac{n}{N_w - 1}\right)\right), 0 \leq n \leq N_w - 1 \quad (4.7)$$

zum Einsatz. Da der Layer I eine Frequenzauflösung von $N_{DFT} = 512$ vorsieht, besitzt das Fenster ebenso die Länge $N_w=512$. Abbildung 4.7 a) zeigt das Hann-Fenster im Vergleich mit einem Rechteck-Fenster. Abbildung 4.7 b) zeigt die Fourier-transformierten Fenster. Das Hann-Fenster besitzt logarithmiert eine Grunddämpfung von 6 dB. Diese kann durch einen Vorfaktor von 2 im Zeitbereich herausgerechnet werden.

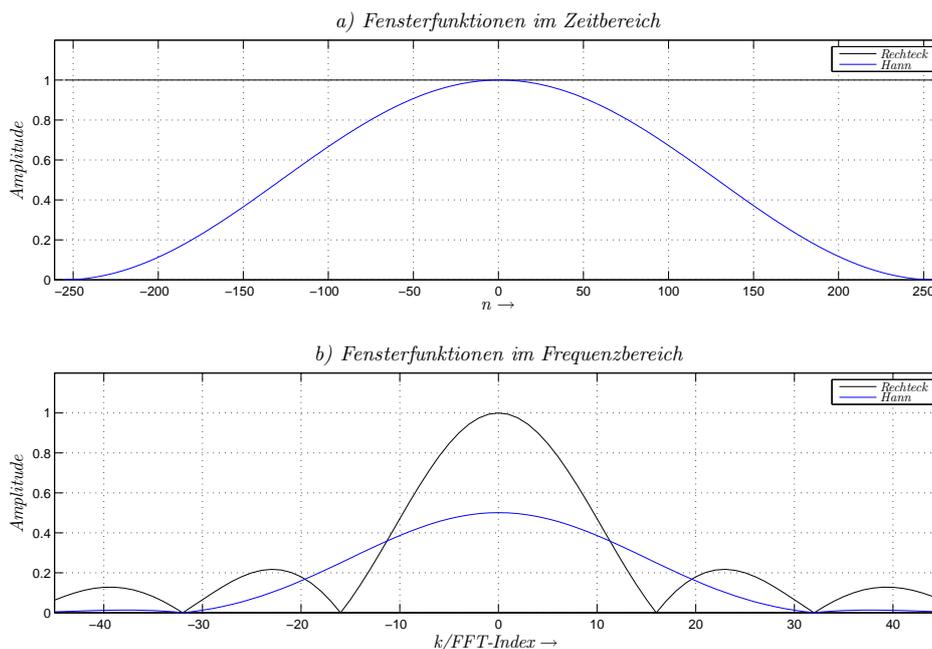


Abbildung 4.7.: Vergleich einer Rechteck- mit einer Hann-Fensterfunktion

⁹Der Leck-Effekt beschreibt das Phänomen bei der Spektralanalyse, dass Frequenzanteile vorkommen, die bei einer theoretisch unendlich langen Beobachtungszeitfolge nicht enthalten sind. Dabei ist die Auswirkung ein in gewissen Maße verschmiertes Frequenzspektrum.

Anschließend wird das LDS durch die DCT

$$\underline{X}(k) = \sum_{n=0}^{N_{DFT}-1} x[n] \cdot 2 \cdot w_{hann}[n] \cdot e^{-j2\pi nk/N_{DFT}} \quad (4.8)$$

gebildet, mit $0 \leq k < (N_{DFT} = 512)$. Für den Betrag gilt

$$X(k) = |\underline{X}(k)| = \sqrt{\underline{X}_{Re}^2(k) + \underline{X}_{Im}^2(k)}. \quad (4.9)$$

Für weitere Berechnungen wird neben Gleichung 4.9 der Leistungspegel

$$X_{dB}(k)/dB = 20 \cdot \log_{10}(X(k)) \quad (4.10)$$

verwendet. Das LDS beruht auf 512 Abtastwerten. Die vorher beschriebenen Skalenfaktoren, welche aus der Filterbank hervorgehen, werden aus insgesamt $12 \cdot 32 = 384$ Abtastwerten ermittelt. Das folgend beschriebene psychoakustische Modell arbeitet mit diesen Werten. Damit die Skalenfaktoren und das LDS Basis selber Abtastwerte sind, bedarf es eines Overlapping der DFT von 128 Abtastwerten.

4.4. Psychoakustisches Modell

Das MPEG-1 Verfahren besitzt zwei unterschiedliche Modelle, die die in Kapitel 2 beschriebenen Maskierungseffekte ausnutzen. Für die Layer I und II wird das psychoakustische Modell 1 verwendet, welches durch das Ablaufdiagramm in Abbildung 4.8 beschrieben wird. Es arbeitet mit dem LDS $X(k)$ und den aus der Filterbank ermittelten Skalenfaktoren $SCF(i)$.

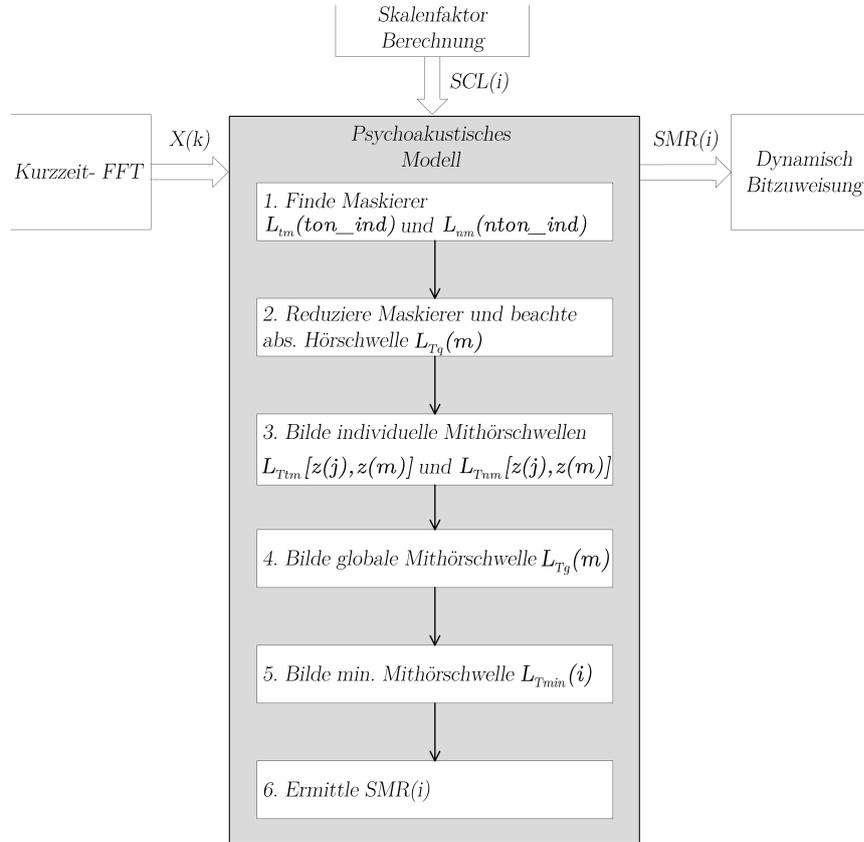


Abbildung 4.8.: Ablaufdiagramm des psychoakustischen Modells 1

Dabei werden die Werte der LUT¹⁰ aus Tabelle 4.4 verwendet. Diese beinhaltet unter anderem die absolute Hörschwelle $L_{T_q}(m)$, mit $0 \leq m < 102$ Werten. Dabei basiert der Index m auf einer reduzierten Anzahl der N_{DFT} relevanten Indices k . Bezogen auf die Spektrallinien des LDS resultiert eine Dezimation dieser, sodass die niederfrequenten Anteile eine höhere Auflösung erhalten als höherfrequente Anteile. Abbildung 4.9 veranschaulicht dies. Aufgrund von Symmetrieeigenschaften der DFT sind nur die Spektrallinien an den Stellen $0 \leq k < (N_{DFT}/2 = 256)$, von $X(k)$, relevant. Das ISO MPEG-1 Layer I mit einer Abtastrate von 48 kHz nutzt dabei einen Frequenzbereich von $0 \text{ Hz} \leq f \leq 20250 \text{ Hz}$, da das menschliche Gehör Anteile darüber hinaus nicht mehr wahrnimmt. Daraus folgt, dass nur Indices bis zu

$$k = \frac{20250 \text{ Hz}}{48000 \text{ Hz}} \cdot 512 = 216 \quad (4.11)$$

benötigt werden.

¹⁰In einer Lookup Tabelle (LUT) werden Werte zur Vermeidung aufwändiger Berechnungen statisch definiert.

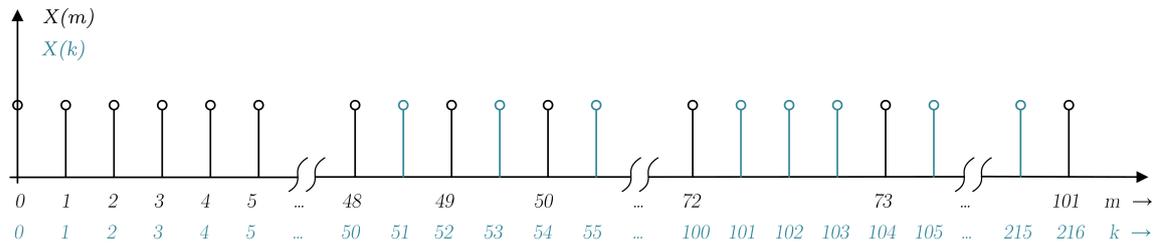


Abbildung 4.9.: Schematische Darstellung des Zusammenhangs von Indices k und m . Alle Funktionen in m beinhalten dabei eine zu hohen Frequenzen hin schrittweise dezimierte Darstellung als Funktionen in k .

Index m	Index k	Freq. in Hz	z in Bark	L_{T_q} in dB	Index m	Index k	Freq. in Hz	z in Bark	L_{T_q} in dB
0	0	93,75	0,925	24,17	51	56	5250,00	18,823	1,06
1	1	187,50	1,842	13,87	52	58	5437,50	19,028	1,39
2	2	281,25	2,742	10,01	53	60	5625,00	19,226	1,66
3	3	375,00	3,618	7,94	54	62	5812,50	19,419	1,88
4	4	468,75	4,463	6,62	55	64	6000,00	19,606	2,08
5	5	562,50	5,272	5,70	56	66	6187,50	19,788	2,27
6	6	656,25	6,041	5,00	57	68	6375,00	19,964	2,46
7	7	750,00	6,770	4,45	58	70	6562,50	20,135	2,65
8	8	843,75	7,457	4,00	59	72	6750,00	20,300	2,86
9	9	937,50	8,103	3,61	60	74	6937,50	20,461	3,09
10	10	1031,25	8,708	3,26	61	76	7125,00	20,616	3,33
11	11	1125,00	9,275	2,93	62	78	7312,50	20,766	3,60
12	12	1218,75	9,805	2,63	63	80	7500,00	20,912	3,89
13	13	1312,50	10,301	2,32	64	82	7687,50	21,052	4,20
14	14	1406,25	10,765	2,02	65	84	7875,00	21,188	4,54
15	15	1500,00	11,199	1,71	66	86	8062,50	21,318	4,91
16	16	1593,75	11,606	1,38	67	88	8250,00	21,445	5,31
17	17	1687,50	11,988	1,04	68	90	8437,50	21,567	5,73
18	18	1781,25	12,347	0,67	69	92	8625,00	21,684	6,18
19	19	1875,00	12,684	0,29	70	94	8812,50	21,797	6,67
20	20	1968,75	13,002	-0,11	71	96	9000,00	21,906	7,19
21	21	2062,50	13,302	-0,54	72	100	9375,00	22,113	8,33
22	22	2156,25	13,586	-0,97	73	104	9750,00	22,304	9,63
23	23	2250,00	13,855	-1,43	74	108	10125,00	22,482	11,08
24	24	2343,75	14,111	-1,88	75	112	10500,00	22,646	12,71
25	25	2437,50	14,354	-2,34	76	116	10875,00	22,799	14,53
26	26	2531,25	14,585	-2,79	77	120	11250,00	22,941	16,54
27	27	2625,00	14,807	-3,22	78	124	11625,00	23,072	18,77
28	28	2718,75	15,018	-3,62	79	128	12000,00	23,195	21,23
29	29	2812,50	15,221	-3,98	80	132	12375,00	23,309	23,94
30	30	2906,25	15,415	-4,30	81	136	12750,00	23,415	26,90
31	31	3000,00	15,602	-4,57	82	140	13125,00	23,515	30,14
32	32	3093,75	15,783	-4,77	83	144	13500,00	23,607	33,67
33	33	3187,50	15,956	-4,91	84	148	13875,00	23,694	37,51

34	34	3281,25	16,124	-4,98	85	152	14250,00	23,775	41,67
35	35	3375,00	16,287	-4,97	86	156	14625,00	23,852	46,17
36	36	3468,75	16,445	-4,90	87	160	15000,00	23,923	51,04
37	37	3562,50	16,598	-4,76	88	164	15375,00	23,991	56,29
38	38	3656,25	16,746	-4,55	89	168	15750,00	24,054	61,94
39	39	3750,00	16,891	-4,29	90	172	16125,00	24,114	68,00
40	40	3843,75	17,032	-3,99	91	176	16500,00	24,171	68,00
41	41	3937,50	17,169	-3,64	92	180	16875,00	24,224	68,00
42	42	4031,25	17,303	-3,26	93	184	17250,00	24,275	68,00
43	43	4125,00	17,434	-2,86	94	188	17625,00	24,322	68,00
44	44	4218,75	17,563	-2,45	95	192	18000,00	24,368	68,00
45	45	4312,50	17,688	-2,04	96	196	18375,00	24,411	68,00
46	46	4406,25	17,811	-1,63	97	200	18750,00	24,452	68,00
47	47	4500,00	17,932	-1,24	98	204	19125,00	24,491	68,00
48	50	4687,50	18,166	-0,51	99	208	19500,00	24,528	68,00
49	52	4875,00	18,392	0,12	100	212	19875,00	24,564	68,00
50	54	5062,50	18,611	0,64	101	216	20250,00	24,597	68,00

Tabelle 4.3.: Indices und korrespondierende Frequenzen sowie Werte der absoluten Hörschwelle und Tonheiten in Bark

Das Ergebnis des psychoakustischen Modells ist ein Signal-Mithörschwellenabstand (*engl. Signal-to-Mask Ratio (SMR)*), welcher eine Größe für den Quantisierungsgrad liefert. Abbildung 4.10 veranschaulicht das Verfahren anhand eines tonalen Schalls. Tritt dieser maskierende Ton L_{tm} in einem Frequenzband auf, so besitzt er einen mit b_{Bit} -Bit codierten SNR und bildet eine Mithörschwelle $L_{T_{tm}}$. Diese erstreckt sich dabei auch über benachbarte Frequenzgruppen. Alle Schalle, die sich unterhalb der Mithörschwelle befinden, können vom menschlichen Gehör nicht erfasst werden. Hieraus geht der Mithörschwellen-Rauschabstand (*engl. Mask-to-Noise Ratio (MNR)*) hervor, welcher vom SNR entlang der Mithörschwelle $L_{T_{tm}}$ verläuft. Der Signal-Mithörschwellenabstand resultiert aus der Differenz von SNR und MNR. Dadurch dass die Rauschgrenze angehoben wird bzw. Signale für das menschliche Gehör im Rauschen untergehen, ist es nicht mehr nötig mit vollen b_{Bit} -Bit zu codieren und es kann eine Bit-Reduktion stattfinden.

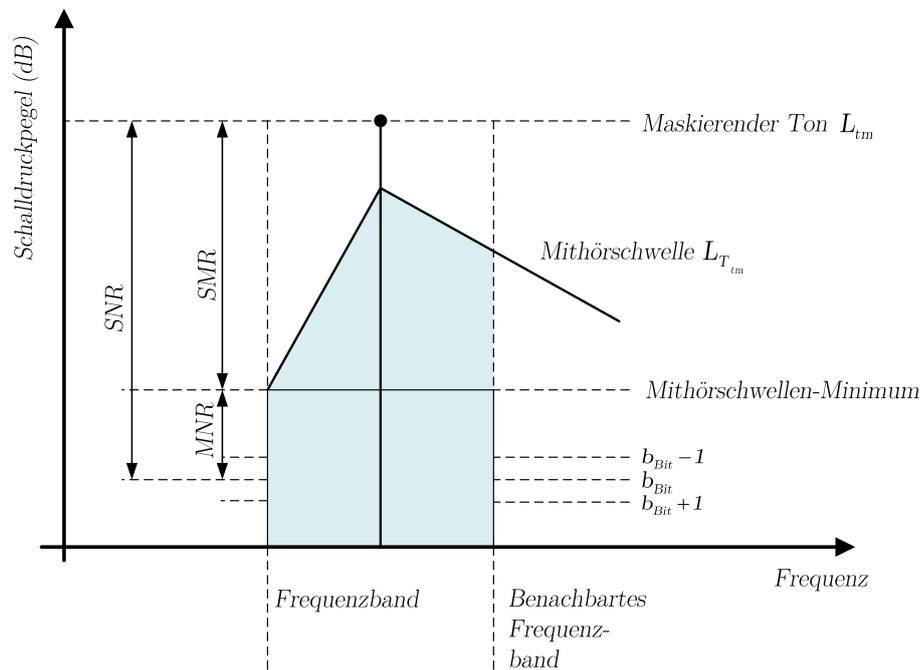


Abbildung 4.10.: Schematische Darstellung der simultanen Maskierung

Das Psychoakustische Modell 1 ermittelt den SMR nicht direkt aus SNR und MNR, sondern korrespondierend dazu aus der Differenz von Schalldruckpegel $L_S(i)$ und einer minimalen Mithörschwelle $L_{T_{min}}(i)$ im Teilband i . Die folgenden 6 Schritte erläutern, wie das Modell anhand tonaler und rauschartiger Maskierer und dessen individueller Mithörschwellen den Signal-Mithörschwellenabstand ermittelt.

1. Finde Maskierer:

Eine Spektrallinie von $X(k)$ ist ein tonaler Maskierer, wenn

$$X(k) > X(k-1) \text{ und } X(k) \geq X(k+1) \quad (4.12)$$

erfüllt ist. Dies kennzeichnet ein lokales Maximum. Des Weiteren muss ein Pegelabstand zu den direkt benachbarten Spektrallinien von

$$X(k) - X(k+j) \geq 7 \text{ dB} \quad (4.13)$$

erfüllt sein. Dabei gilt:

$$j = \begin{cases} [-2; 2] & , \text{ für } 2 \leq k < 63 \\ [-3; -2; 2; 3] & , \text{ für } 63 \leq k < 127 \\ [-6; -5; -4; -3; -2; 2; 3; 4; 5; 6] & , \text{ für } 127 \leq k \leq 250 \end{cases} \quad (4.14)$$

Dabei kann es auch dazu kommen, dass kein tonaler Maskierer gefunden wird. Ist jedoch mindestens einer vorhanden, so wird der Schalldruckpegel des *ton_ind*-ten Maskierers an der Stelle k durch die Summe des entlogarithmierten Pegels sowie anhand dessen direkt benachbarten gemäß

$$L_{tm}(ton_ind) = 10 \cdot \log_{10} \left(10^{\frac{X(k-1)}{10}} + 10^{\frac{X(k)}{10}} + 10^{\frac{X(k+1)}{10}} \right) \quad (4.15)$$

gebildet. Im Gegensatz zu tonalen Maskierern kommt ein rauschartiger bzw. nicht tonaler Maskierer in jedem Bark-Frequenzband vor. Dabei befinden sich diese an der Stelle des geometrischen Mittels der Frequenzindices k im jeweiligen Frequenzband z in Bark.

$$\begin{aligned} \bar{k} &= \left(\prod_{j=l_B}^{u_B} k \right)^{\frac{1}{l_B - u_B + 1}} \\ \bar{k} &= [0; 1; 2; 3; 4; 5; 6; 8; 9; 11; 13; 16; 18; \dots \\ & 21; 25; 30; 36; 44; 51; 61; 73; 89; 107; 139; 183] \end{aligned} \quad (4.16)$$

mit dem unteren Frequenzband-Index l_B und u_B für den oberen Frequenzband-Index. Der Leistungspegel nicht-tonaler Maskierer an der Stelle \bar{k} ergibt sich zu

$$L_{nm}(nton_ind) = 10 \cdot \log_{10} \left(\sum_{j=l_B}^{u_B} 10^{\frac{X(j)}{10}} \right). \quad (4.17)$$

2. Reduziere Maskierer und beachte abs. Hörschwelle:

Der nächste Schritt ist eine Reduzierung der Maskierer, wenn diese einen geringeren Pegel als die absolute Hörschwelle $L_{T_q}(m)$ besitzen. Des Weiteren kann der schwächere zweier tonaler Maskierer wegfallen, wenn diese sich näher als $\Delta z = 0,5$ Bark sind.

3. Bilde individuelle Mithörschwellen:

Die Mithörschwellen werden gemäß

$$L_{T_{tm}}[z(j), z(m)]/\text{dB} = L_{tm}[z(j)] + av_{tm}[z(j)] + v_f[z(j), z(m)] \quad (4.18)$$

für tonale Maskierer und

$$L_{T_{nm}}[z(j), z(m)]/\text{dB} = L_{nm}[z(j)] + av_{nm}[z(j)] + v_f[z(j), z(m)] \quad (4.19)$$

für nicht-tonale Maskierer berechnet. Dabei ist

$$av_{tm}/\text{dB} = -1,525 - 0,275 \cdot z(j) - 4,5 \quad (4.20)$$

der Maskierungsindex für tonale und

$$av_{nm}/\text{dB} = -1,525 - 0,175 \cdot z(j) - 0,5 \quad (4.21)$$

der Maskierungsindex für nicht-tonale Maskierer. Dieser ist ein Barkband-abhängiger Offset des Leistungspegels. Die Maskierungsfunktion bzw. Mithörschwelle bildet $v_f[z(j), z(m)]$, mit dem Frequenzgruppenabstand $\Delta z = z(m) - z(j)$, mit z in Bark.

$$v_f[z(j), z(m)]/\text{dB} = \begin{cases} 17 \cdot (\Delta z + 1) + (0,4 \cdot X[z(j)] + 6) & , -3 \leq \Delta z < -1 \\ (0,4 \cdot X[z(j)] + 6) \cdot \Delta z & , -1 \leq \Delta z < 0 \\ -17 \cdot \Delta z & , 0 \leq \Delta z < 1 \\ -(\Delta z - 1) \cdot (17 - 0,15 \cdot X[z(j)]) - 17 & , 1 \leq \Delta z < 8 \end{cases} \quad (4.22)$$

Der Index j steht für die Stelle der Spektrallinie, an dem sich der Maskierer L_{tm} oder L_{nm} befindet. Die Maskierungsfunktion v_f beschreibt die Maskierung in der Frequenzgruppe $z(m)$ durch den Maskierer in der Frequenzgruppe $z(j)$. Aus Gleichung 4.22 geht hervor, dass eine Mithörschwelle benachbarte Frequenzgruppen im Bereich $-3 \leq z \leq 7$, demnach über 10 Bark, mit einbezieht.

Die folgende Simulation veranschaulicht die Prozedur. Als Test-Signal wird eine Überlagerung von drei Sinussignalen mit etwas Rauschen verwendet.

$$\begin{aligned}
 x(t) = & \sin(2\pi \cdot 500 \text{ Hz} \cdot t) \\
 & + 0,25 \cdot \sin(2\pi \cdot 1500 \text{ Hz} \cdot t) \\
 & + 0,50 \cdot \sin(2\pi \cdot 7500 \text{ Hz} \cdot t) \\
 & + 3 \cdot 10^{-3} \cdot \text{noise}(t)
 \end{aligned} \tag{4.23}$$

Abbildung 4.11 a) zeigt das Zeitsignal $x(t)$, Abbildung 4.11 b) das auf 96 dB normierte LDS $X(k)$ bzw. $|X(e^{j2\pi f})|$.

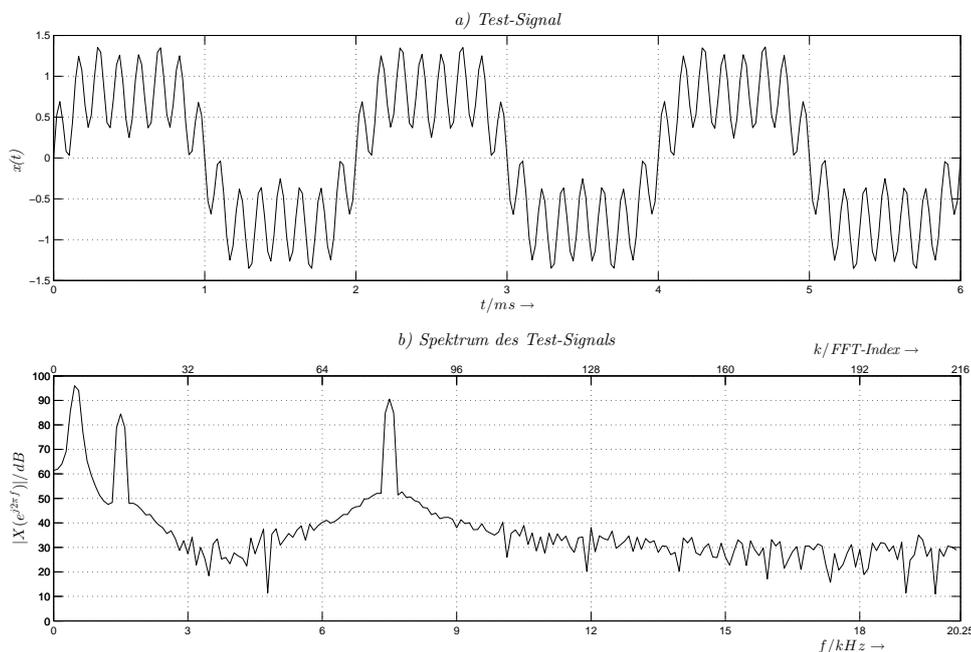


Abbildung 4.11.: Für Simulationszwecke verwendetes Test-Signal

Aus dem Test-Signal können eindeutig drei tonale Maskierer ermittelt werden, wovon jeder eine Mithörschwelle bildet, vgl. Abbildung 4.12 a). Abbildung 4.12 b) zeigt die Maskierer und Mithörschwellen der nicht-tonalen Anteile. Es wird deutlich, dass tiefere Frequenzen eine höhere Auflösung besitzen. Dies ist der Einfluss der Frequenzgruppen. Abbildung 4.13 a) und b) zeigen die korrespondierende Skalierung über einer Bark-Skala. Anhand der Frequenzgruppen ist der Bereich von 10 Bark zu erkennen, über den sich eine Mithörschwelle erstreckt. Die obere und untere Flankensteilheit besitzt Ähnlichkeit mit der von Zwicker aufgestellten Gleichung 2.6 und Gleichung 2.7 bzw. der Mithörschwellen in Abbildung 2.6 des vorangegangenen Kapitels 2 der Psychoakustik.

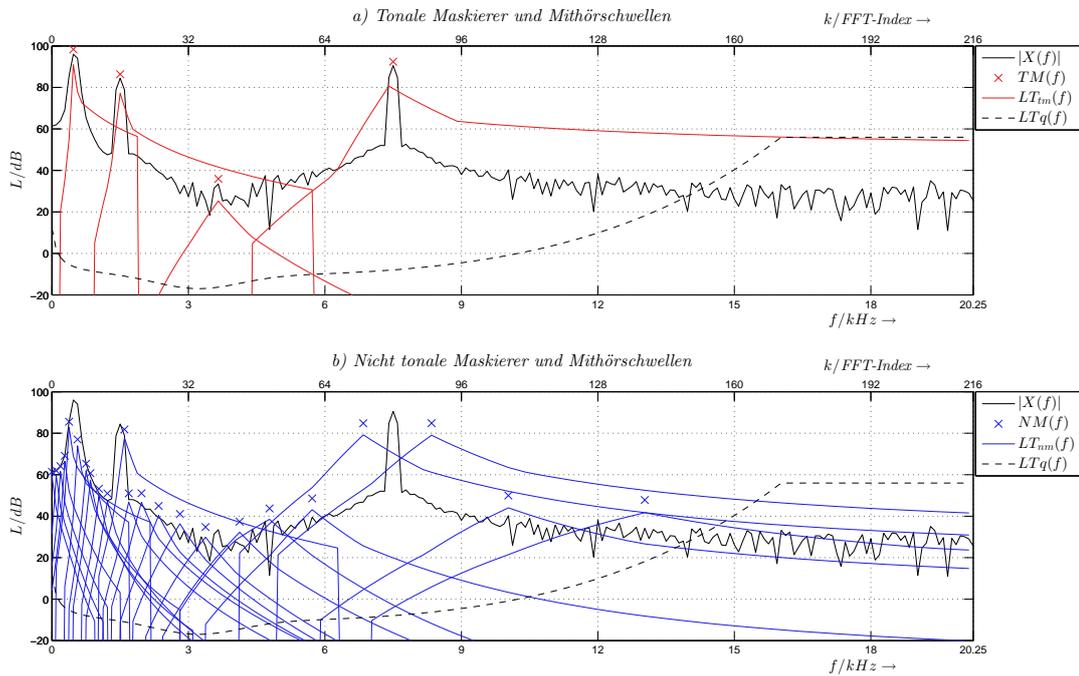


Abbildung 4.12.: Maskierer und Mithörschwellen resultierend aus dem Test-Signal, aufgetragen auf der Frequenz-Skala

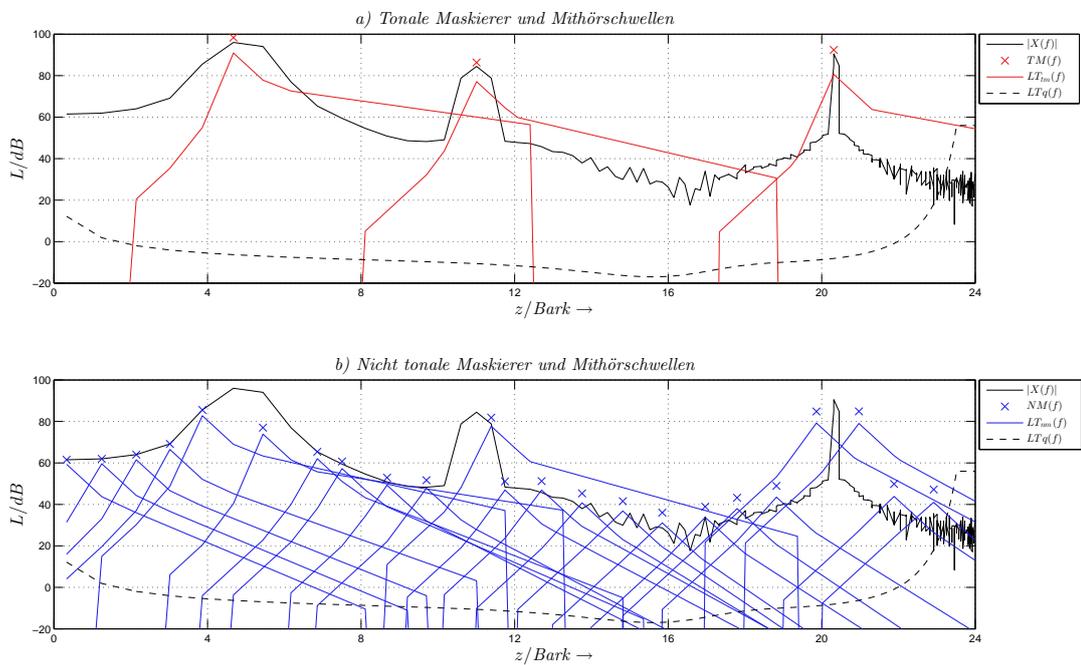


Abbildung 4.13.: Maskierer und Mithörschwellen resultierend aus dem Test-Signal, aufgetragen auf der Bark-Skala

4. Bilde globale Mithörschwelle:

Die globale Mithörschwelle bildet sich aus der Überlagerung aller entlogarithmierten tonalen und nicht-tonalen Mithörschwellen sowie der absoluten Hörschwelle.

$$L_{T_g}(m)/\text{dB} = 10 \cdot \log_{10} \left(10^{L_{T_q}(m)} + \sum_{j=1}^{T_m} 10^{L_{T_{tm}}[z(j),z(m)]} + \sum_{j=1}^{N_m} 10^{L_{T_{nm}}[z(j),z(m)]} \right), \quad (4.24)$$

dabei ist T_m die Anzahl der tonalen und N_m die Anzahl der nicht-tonalen Maskierer. Abbildung 4.14 zeigt die globale Mithörschwelle nach Gleichung 4.24, gebildet aus den Mithörschwellen in Abbildung 4.12 bzw. Abbildung 4.13.

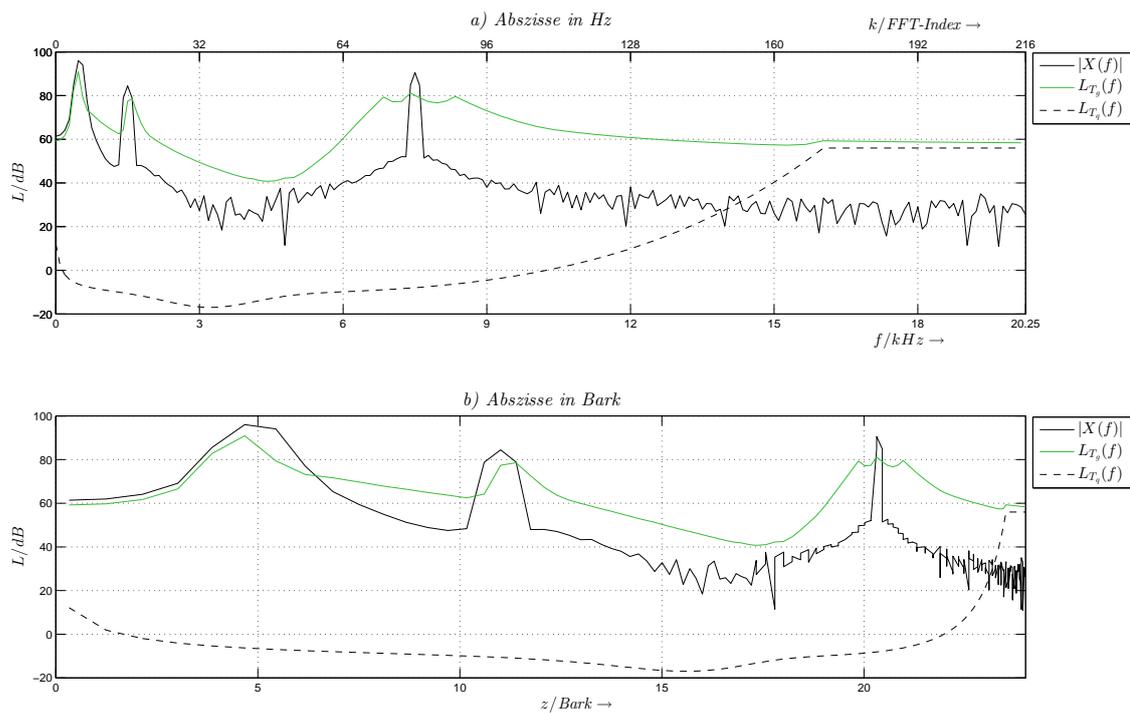


Abbildung 4.14.: Globale Mithörschwelle $L_{T_g}(f)$ resultierend aus dem Test-Signal

5. Bilde minimale Mithörschwelle:

Die minimale Mithörschwelle entspricht dem kleinsten Mithörschwellen-Rauschabstand der globalen Mithörschwelle $L_{T_g}(m)$, welcher im Filterbank-Teilband i vorkommt.

$$L_{T_{min}}(i)/\text{dB} = \min [L_{T_g}(m)], \text{ mit } 0 \leq m < 102 \quad (4.25)$$

Hier muss eine Umrechnung bzw. Mapping der Indices m in die 32 Teilbänder geschehen.

6. Ermittle den Signal- Mithörschwellenabstand:

Der Signal-Mithörschwellenabstand ergibt sich schließlich zu

$$SMR(i)/\text{dB} = L_S(i) - L_{T_{min}}(i), \text{ mit } 0 \leq i < 32. \quad (4.26)$$

Dabei ist der Schalldruckpegel gemäß

$$L_S(i)/\text{dB} = \max [X_{max}(i), 20 \cdot \log_{10}(SCF(i) \cdot 2^{15}) - 10 \text{ dB}] \quad (4.27)$$

aus dem Maximum der im Teilband i vorkommenden Spektrallinien $X_{max}(i)$ und den logarithmierten Skalenfaktoren $SCF(i)$ ermittelt. Das LDS $X(k)$ ist dabei auf 96 dB normiert, welches ungefähr dem SNR für $b_{Bit}=16$ Bit entspricht. Durch die Multiplikation von 2^{15} mit den Skalenfaktoren werden auch diese auf 96 dB normiert, da der größtmögliche Wert 2 ist (vgl. Tabelle 4.1 der Skalenfaktoren). Der Schalldruckpegel $L_s(i)$ und die minimale Mithörschwelle $L_{T_{min}}(i)$ des Test-Signals sind in Abbildung 4.15 a) über die 32 Teilbänder dargestellt. Der nach Gleichung 4.26 berechnete Signal-Mithörschwellenabstand ist in Abbildung 4.15 b) abgebildet. Unterschreitet dieser in einem Teilband die Grenze von 0 dB, so bedeutet dies, dass die Mithörschwelle über dem Schallpegel liegt. Schalle sind in diesem Bereich für unser Gehör weitestgehend maskiert und daher wenig bis gar nicht hörbar. Demnach bedarf es auch nur wenig bis gar keine Codierung bzw. Datenübertragen dieser Signalanteile. Jene Teilbänder, die einen SMR über 0 dB besitzen, werden am meisten Bits zugewiesen.

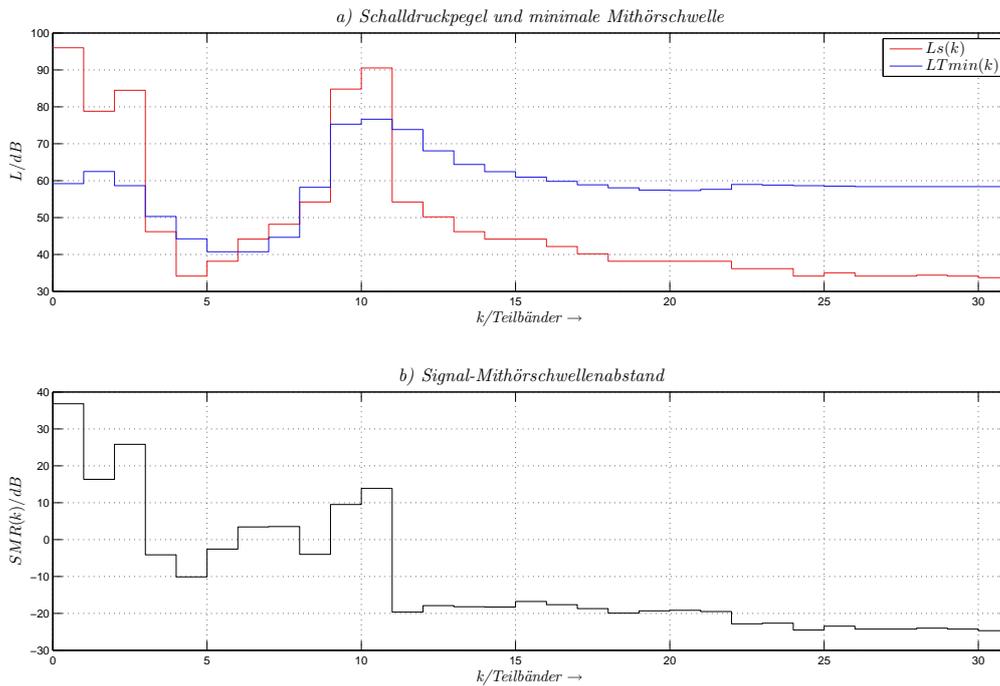


Abbildung 4.15.: Ermittlung des Signal- Mithörschwellenabstands aus Schalldruckpegel und minimaler Mithörschwelle

4.5. Dynamische Bitzuweisung

Die Bitzuweisung geschieht für jedes der 32 Teilbänder. Wird die Bitanzahl gesteigert, so erhöht sich der MNR. Dies wird auch durch am Anfang des Kapitels aufgezeigte Abbildung 4.10 verdeutlicht. Wird b_{Bit} um ein Bit inkrementiert, so wird der Mithörschwellen-Rauschabstand größer. Der MPEG-1 Standard erlaubt ein Maximum von 15 Bit pro Teilband, woraus mit Gleichung 4.4 ein SNR von 92 dB resultiert. Wie viel Bit schließlich für ein Teilband zugewiesen wird, kennzeichnet $BSPL(i)$ und muss bei der Übertragung in einem Rahmen hinterlegt sein. Hierfür reicht eine 4 Bit Darstellung aus. Tabelle 4.4 zeigt den erreichbaren SNR in Abhängigkeit der Bitanzahl sowie den 4-Bit Code $BSPL$. Wie viele Bits für einen Rahmen verfügbar sind, wird durch die Bitrate bestimmt. Das Ablaufdiagramm für die Bitzuweisung ist in Abbildung 4.16 dargestellt. Für die folgende Beschreibung wird von der Bitrate $v_B=192$ kbps ausgegangen, welches einem Kompressionsgrad von 1:4 gleich kommt.

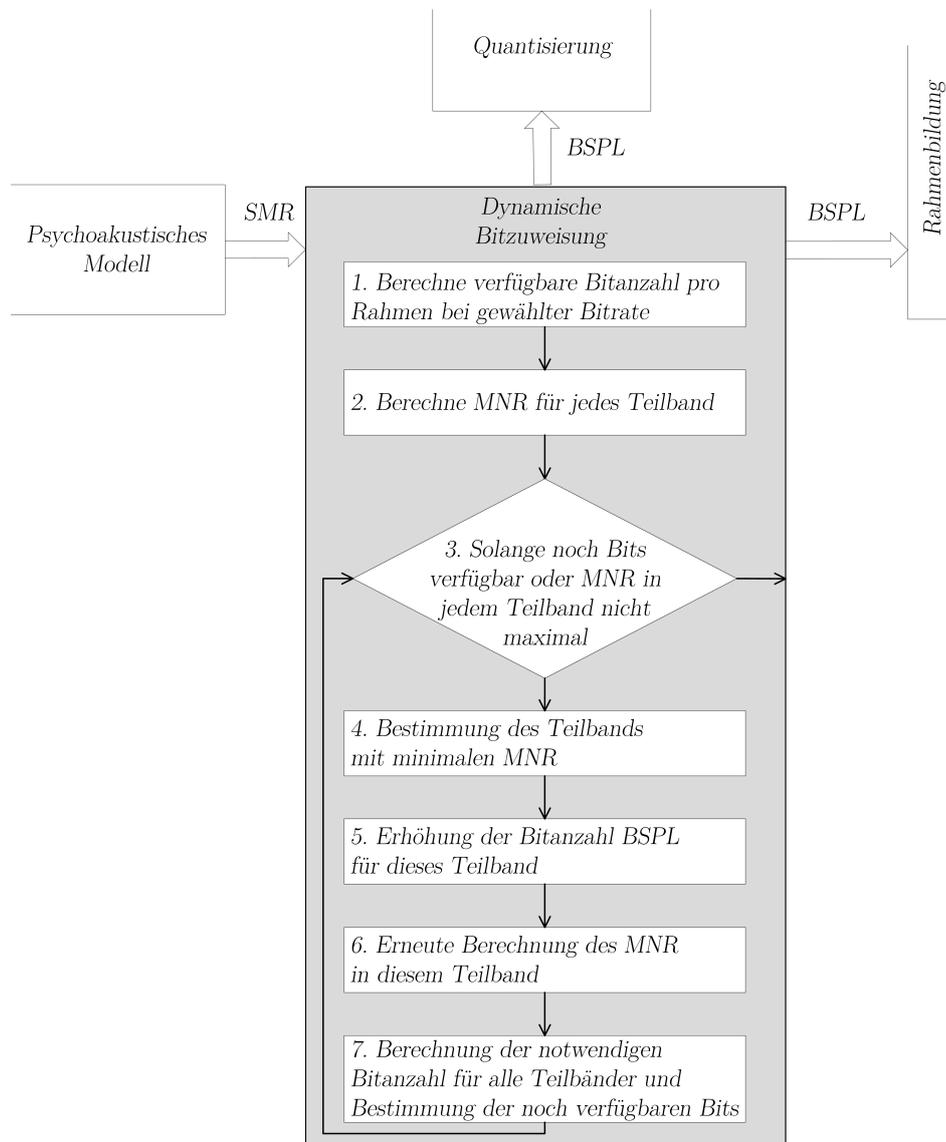


Abbildung 4.16.: Ablaufdiagramm der dynamischen Bitzuweisung

Zugewiesene Bits	BSPL Code	SNR/dB
0	0000	0,00
2	0001	7,00
3	0010	16,00
4	0011	25,28
5	0100	31,59
6	0101	37,75
7	0110	43,84
8	0111	49,89
9	1000	55,93
10	1001	61,96
11	1010	67,98
12	1011	74,01
13	1100	80,03
14	1101	86,05
15	1110	92,01
unzulässig	1111	

Tabelle 4.4.: Bitzuweisung

1. Berechne verfügbare Bitanzahl pro Rahmen bei gewählter Bitrate:

Die gesamte verfügbare Anzahl an Bits pro Rahmen ist gegeben durch

$$cb/\text{Bit} = \frac{12 \cdot 32 \cdot 16 \text{ Bit}}{cr}, \text{ mit } cr = \frac{48 \text{ kHz} \cdot 16 \text{ Bit}}{v_B}. \quad (4.28)$$

Gleichung 4.28 ergibt sich aus den 32 Teilfilterbändern mit je 12 Abtastwerten und einer 16 Bit Auflösung, welche mit einem Kompressionsgrad von cr pro Frame codiert werden sollen. Für $v_B = 192$ kbps ergibt sich eine Rahmengröße von $cb = 1536$ Bit.

2. Berechne MNR für jedes Teilband:

Um die maximale Bitanzahl bei gegebener Bitrate zuweisen zu können, wird der Mithörschwellen-Rauschabstand

$$MNR(i) = SNR(i) - SMR(i) \quad (4.29)$$

benötigt. Der Signal-Mithörschwellenabstand $SMR(i)$ ist das Ergebnis des psychoakustischen Modells. Der $SNR(i)$ entstammt aus der Tabelle 4.4. Dieser wird gewählt, dass der Mithörschwellen-Rauschabstand, Gleichung 4.29, größer gleich Null ist. Im nächsten Schritt beginnt ein iterativer Prozess. Dieser ermittelt die maximal

mögliche Bitanzahl $BSPL(i)$ pro Teilband, wodurch der $MNR(i)$ bzgl. der Bitrate dynamisch optimiert wird.

3. Solange noch Bits verfügbar oder MNR in jedem Teilband nicht maximal:

Die noch verfügbare Bitanzahl ist gegeben durch

$$adb/\text{Bit} = cb - (bn + bbal + banc), \quad (4.30)$$

wobei $bbal$ für die mit 4-Bit codierte Bitanzahl $BSPL(i)$ (aus Tabelle 4.4) der Bitzuweisung je Teilband steht. Dieser ist statisch und beträgt $4 \cdot 32 = 128$ Bit. $banc$ ist eine Bitanzahl pro Frame, welche für Zusatzinformationen reserviert ist. Diese ist optional und wird zu Null angenommen. Die im laufenden Prozess zugewiesenen Bits sind mit bn gekennzeichnet. Die Bitzuweisung ist beendet und die Prozess-Schleife wird verlassen, falls in jedem Teilband der $MNR(i)$ maximal ist (das bedeutet, es wurden 15 Bits zugewiesen oder der $SNR(i)$ ist maximal) oder die verfügbaren Bits, Gleichung 4.30, sind weniger als $(2 \cdot 12 + 6) = 30$. Dies ist die höchste Anzahl an Bits, die in einer Prozess-Schleife zugewiesen werden können. Bei einer erstmaligen Zuweisung werden 2 Bit für 12 Abtastwerte und 6 Bit für den zugehörigen Skalenfaktor vergeben.

4. Bestimmung des Teilbands mit minimalen MNR:

Zunächst wird das Teilband gesucht, welches den kleinsten Mithörschwellen-Rauschabstand besitzt.

5. Erhöhung der Bitanzahl BSPL für dieses Teilband:

Das gefundene Teilband erhält die nächsthöhere Bitanzahl $BSPL(i)$ und den nächstgrößeren $SNR(i)$ nach Tabelle 4.4. Wurde in diesem Teilband erstmalig die Bitanzahl erhöht, so wird der Skalenfaktor mit übertragen. Dieser ist in Tabelle 4.1 gegeben. Für eine effiziente Bitzuweisung reicht es aus, den Index d des Skalenfaktors zu übertragen. Bei 63 Werten genügen demnach 6 Bit, was durch $BSCF(i)=6$ Bit gekennzeichnet wird.

6. Erneute Berechnung des MNR in diesem Teilband:

Nach Gleichung 4.29 wird der neue Mithörschwellen-Rauschabstand für dieses Teilband berechnet.

7. Berechnung der notwendigen Bitanzahl für alle Teilbänder und Bestimmung der noch verfügbaren Bits:

Die notwendige bzw. bis zum aktuellen Zeitpunkt zugewiesene Bitanzahl ist durch

$$bn = \sum_{i=0}^{32-1} BSCF(i) + 12 \cdot BSPL(i) \quad (4.31)$$

gegeben. Für jedes Teilband werden 12 Abtastwerte mit $BSPL(i)$ Bit quantisiert, dessen Skalenfaktor durch die Anzahl $BSCF(i)$ in Bit beschrieben ist. Im nächsten Schritt wird die Iterationsbedingung aus Punkt 3 geprüft und ggf. die Bitzuweisung fortgeführt.

Zur Veranschaulichung wird die Bitzuweisung für das Test-Signal aus Abbildung 4.11 bzw. Gleichung 4.23 simuliert. Abbildung 4.17 a) zeigt dabei noch einmal den $SMR(i)$. Die rote Linie in Abbildung 4.17 b) zeigt die minimale Bitanzahl, die bei gegebenem $SMR(i)$ nötig ist. Die grüne Linie zeigt die Bitanzahl nach der dynamischen Zuweisung für die Bitrate $v_B=192$ kbps. In Abbildung 4.17 c) ist der korrespondierende $MNR(i)$ dargestellt.

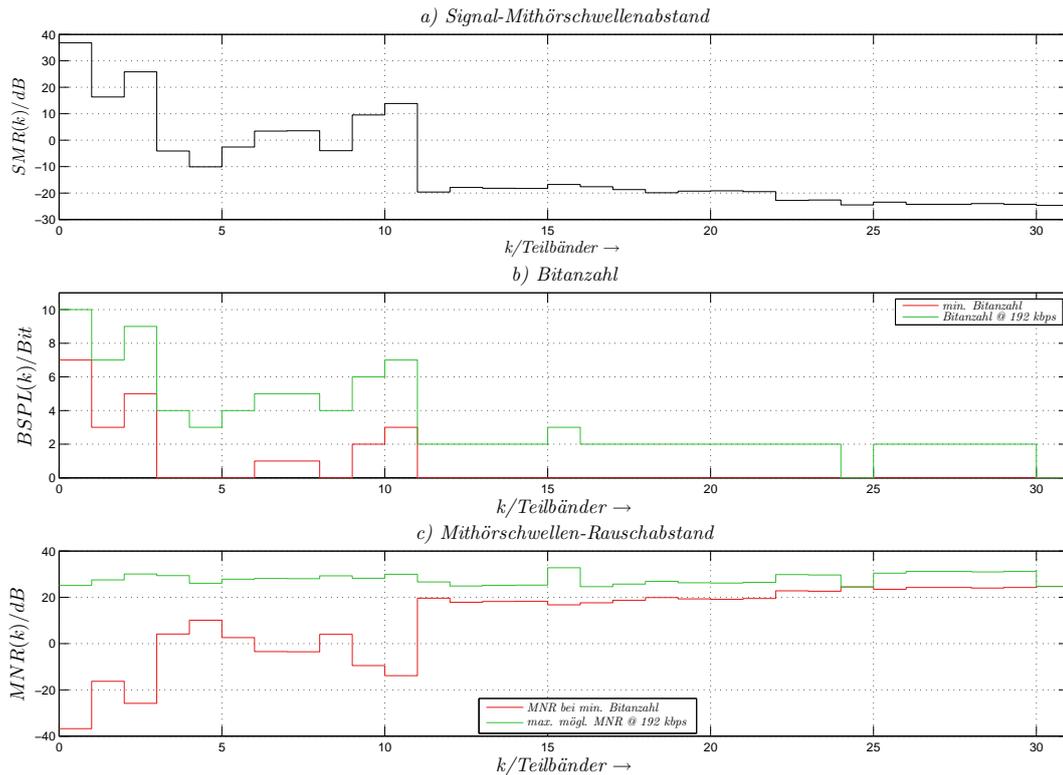


Abbildung 4.17.: Dynamische Bitzuweisung für das Test-Signal bei $v_B=192$ kbps

Die Teilbänder mit einem $SMR(i)$ größer Null besitzen vor der Bitzuweisung den kleinsten $MNR(i)$ und erhalten die größte Anzahl der verfügbaren Bits. Das Ergebnis zeigt einen über alle Frequenzteilbänder nahezu gleichmäßigen $MNR(i)$. Tabelle 4.5 listet die vergebenen Bitanzahlen je Teilband i auf. Nach Gleichung 4.31 werden insgesamt $bn=1404$ Bits zugewiesen. Mit $bbal=128$ Bit bleiben nach Gleichung 4.30 noch $adb=4$ Bits übrig. Für das gegebene Test-Signal ist nach dem psychoakustischen Modell 1 das Signal damit ausreichend codiert.

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
BSPL(i)	10	7	9	4	3	4	5	5	4	6	7	2	2	2	2	2
BSCF(i)	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
BSPL(i)	2	2	2	2	2	2	2	2	2	0	2	2	2	0	2	2
BSCF(i)	6	6	6	6	6	6	6	6	6	0	6	6	6	0	6	6

Tabelle 4.5.: Zugewiesene Bits für die 32 Teilbänder (Simulation)

4.6. Quantisierung

In diesem Abschnitt wird erläutert, wie die 12 mal 32 Teilbandabtastwerte der Analysefilterbank anhand der Daten aus der dynamischen Bitzuweisung ($BSPL(i)$) quantisiert werden. Dies führt unweigerlich zu Verlusten, weshalb auch von einer verlustbehafteten Komprimierung gesprochen wird. Zum Einsatz kommt eine *Mid-Thread* Quantisierungskennlinie, welche durch ihren Nulldurchgang gekennzeichnet ist. Ist der Betrag des Eingangswerts kleiner als $2^{-BSPL(i)}$, so bleibt das Ausgangssignal stets Null. Im Gegensatz zur *Mid-Rise* Kennlinie, welche den Übergangsbereich der Quantisierungsstufe bei Null besitzt. Der Vorteil einer *Mid-Thread* Quantisierungskennlinie ist die Vermeidung, dass bei Rauschen das Ausgangssignal zwischen den ersten Quantisierungsstufen hin und her springt (granulares Rauschen) [Kam11]. Jedes der 12 Abtastwerte in einem Teilband $X(i, s)$ wird durch den Skalenfaktor $SCF(i)$ geteilt und gemäß [Str09]

$$X_Q(i, s) = \operatorname{sgn}\left\{X(i, s)\right\} \cdot \left\lfloor \frac{|X(i, s)|}{SCF(i)} + \frac{1}{2} \right\rfloor, \text{ mit } 0 \leq s < 12 \quad (4.32)$$

auf einen ganzzahligen Wertebereich von $-(2^{BSPL(i)-1}) \leq X_Q(i, s) \leq 2^{BSPL(i)-1} - 1$ gebracht. Abbildung 4.18 zeigt die Quantisierungskennlinie für $BSPL = 3$ Bit.

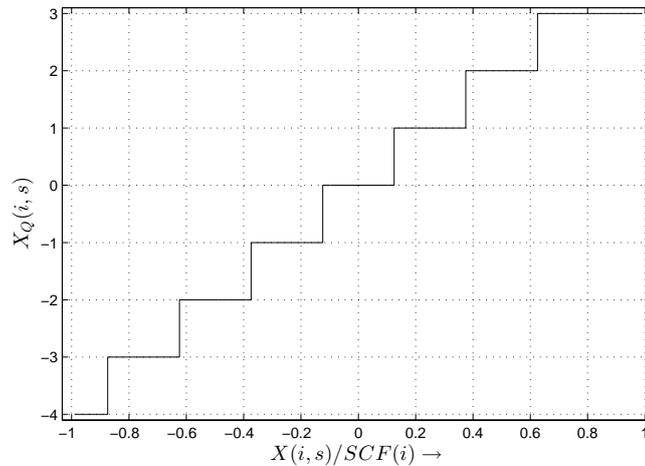


Abbildung 4.18.: Kennlinie einer 3 Bit Quantisierung

4.7. Rahmenbildung

Der Audio-Rahmen, nach dem MPEG-1 Verfahren für alle drei Layer, ist in Abbildung 4.19 dargestellt. In dieser Arbeit wird nur der grün-hinterlegte Teil verwendet, da der Fokus bei den Nutzdaten liegt und weniger bei den Metadaten. Der aus 32 Bit bestehende Header besitzt bspw. Informationen zur Bitrate, zum Layer, ob der Audiostream Copyrights unterliegt und einigen mehr. Unter anderem sind darunter 12 Bits zur Erkennung eines neuen Rahmens enthalten. Diese Synchronisationsfolge wird benötigt, damit der Decoder den Beginn eines Rahmens ermitteln kann.



Abbildung 4.19.: MPEG-1 Audio Rahmen (grün: genutzter Inhalt)

Die Anordnung des Datenfeldes zeigt Abbildung 4.20. Vorne an stehen die Bitanzahlen $BSPL$ jedes Teilband zu je 4 Bit. Im Anschluss folgen die Indices der Skalenfaktoren $BSCF$ mit je 6 Bit. Diese sind nur vorhanden für die Teilbänder, die durch die Bitzuweisung codiert wurden ($BSPL > 0$). Darauf folgen die quantisierten Teilbandabstastwerte $X_Q(i, s)$.



Abbildung 4.20.: Anordnung des Datenfelds

4.8. Decodierung

Der Decoder benötigt zunächst die Information über die Bitanzahl $BSPL(i)$ aus dem Rahmen. Wenn ein Teilband codiert wurde ($BSPL(i) > 0$), wird im Anschluss der Index für den Skalenfaktor $SCF(i)$ dieses Teilbands ausgelesen. Darauf folgt die Dequantisierung der Teilbandsignale $X_Q(i, s)$.

$$X_D(i, s) = \begin{cases} X_Q(i, s) \cdot SCF(i) & , \text{ für } BSPL(i) > 0 \\ 0 & , \text{ für } BSPL(i) = 0 \end{cases} \quad (4.33)$$

Den letzten Schritt bildet die Signalsynthese. Die Rekonstruktion des Audiosignals $\hat{x}[n]$ ergibt sich durch die Filterung der Teilbandabstastwerte mit der Synthese-Filterbank. Abbildung 4.21 a) zeigt das korrespondierende Blockschaltbild mit Aufwärtstastung um $M=32$. Wie auch bei dem Blockschaltbild der Analyse-Filterbank (Abbildung 4.5 b)) zeigt Abbildung 4.21 b) eine für die Realisierung anschaulichere Darstellung mit rotierendem Schalter. Die Koeffizienten der Polyphasenteilfilter $E_{0...63}(-z^2)$ entstammen ebenso dem Prototypenfilter $H_0(z)$ aus Abbildung 4.4.

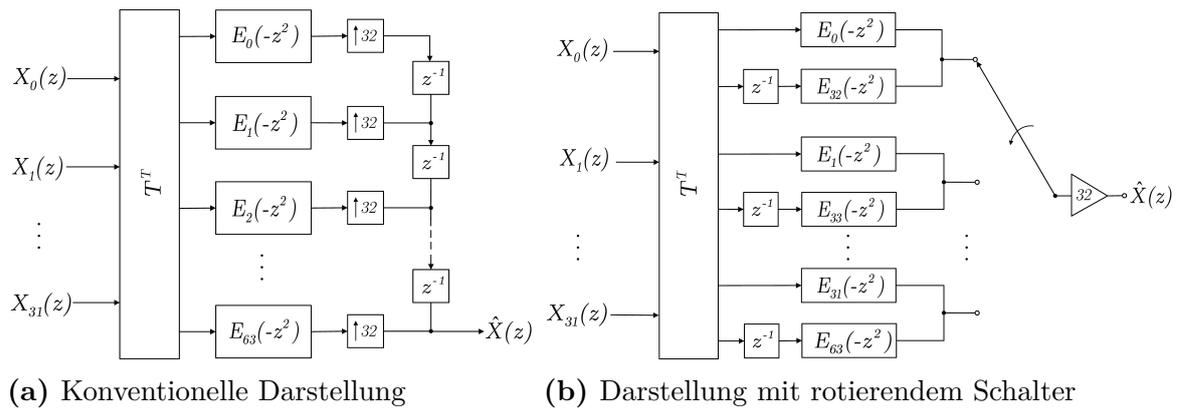


Abbildung 4.21.: Blockschaltbild der Cosinus-modulierten Synthese-Filterbank mit Aufwärtstastung

Das Konzept des MPEG-1 Audio Layer I Verfahrens ist abgeschlossen. Bevor die Entwicklung auf der Hardwareplattform geschieht, werden einige Simulationen durchgeführt, welche später mit den Messergebnissen der Hardwarerealisierung verglichen werden.

4.9. Simulationsergebnisse

Für die Beurteilung des Kompressionsverfahrens und der Analyse unterschiedlicher Bitraten sollen Spektrogramme¹¹ dienen. Zum Einsatz kommen drei unterschiedliche Musikstücke, die in Tabelle 4.6 aufgelistet sind. Dabei wurden verschiedene Musikrichtungen gewählt. Für eine frequenz-selektive Analyse dient zusätzlich ein Wobbelsignal.

Interpret	Musikstück	Musikrichtung	zeitl. Pos.	Bezeichnung
J. Brahms (Dirigent: R. Chailly)	Serenades No. 1, D-Dur, op. 11	Klassik	9:46-10:36	Bra
Johnossi	Roscoe	Alternativ-Rock	1:41-2:35	Joh
P. Kalkbrenner	Sky and Sand	Elektromusik	0:44-1:45	Kal
weitere Audiosignale für Testzwecke				
Wobbelsignal (Sweep)	20 Hz...20 kHz	Testsignal	0:00-0:10	Swe

Tabelle 4.6.: Verwendete Musikstücke für Messungen und Hörtests. Neben dem Namen der Stücke sind Interpret (ggf. Dirigent), Musikrichtung und die zeitliche Position eingetragen. Die Musikstücke sind durch die Notation $\mathcal{S}^{\text{Bezeichnung}}$ gekennzeichnet.

4.9.1. Spektrogramme

Für die Berechnung der Spektrogramme wird eine 512-Punkte FFT mit einem Overlap von 50 % sowie einem Hamming Fenster verwendet. Der betrachtete Frequenzbereich erstreckt sich von 0...20 kHz. Dargestellt wird das Spektrogramm des Musikstücks/Signals $\mathcal{S}^{\text{Bezeichnung}}$ als a) original, b) komprimiert und c) Differenz. Dabei werden die Kompressionsgrade 1:12 und 1:2 bzw. die Bitraten $v_B = 64$ kbps und $v_B = 384$ kbps betrachtet. Ein Vergleich weiterer Bitraten ist im Anhang B gegeben. Abbildung 4.23 zeigt das Wobbelsignal \mathcal{S}^{Swe} bei $v_B = 384$ kbps und Abbildung 4.22 zeigt das Wobbelsignal \mathcal{S}^{Swe} bei $v_B = 64$ kbps. Alle Signalanteile werden über den Frequenzbereich bei beiden Bitraten klar rekonstruiert. Dabei zeigt $v_B = 64$ kbps wie zu erwarten eine stärkere Differenz zwischen Original- und Komprimiertem-Spektrogramm. Auffällig ist das bessere Ergebnis bei tieferen als bei höheren Frequenzen. Dies ist der Einfluss der Bark-Frequenzgruppen nach Zwicker, wonach tiefere Frequenzen für das menschliche Gehör besser wahrgenommen werden können als höhere Frequenzen und somit auch feiner codiert werden müssen.

¹¹Ein Spektrogramm ist die Darstellung eines zeitvarianten Leistungsdichtespektrums. Abgebildet wird es in einem Zeit-Frequenz-Diagramm, wobei die Leistung eine farbliche Skalierung besitzt.

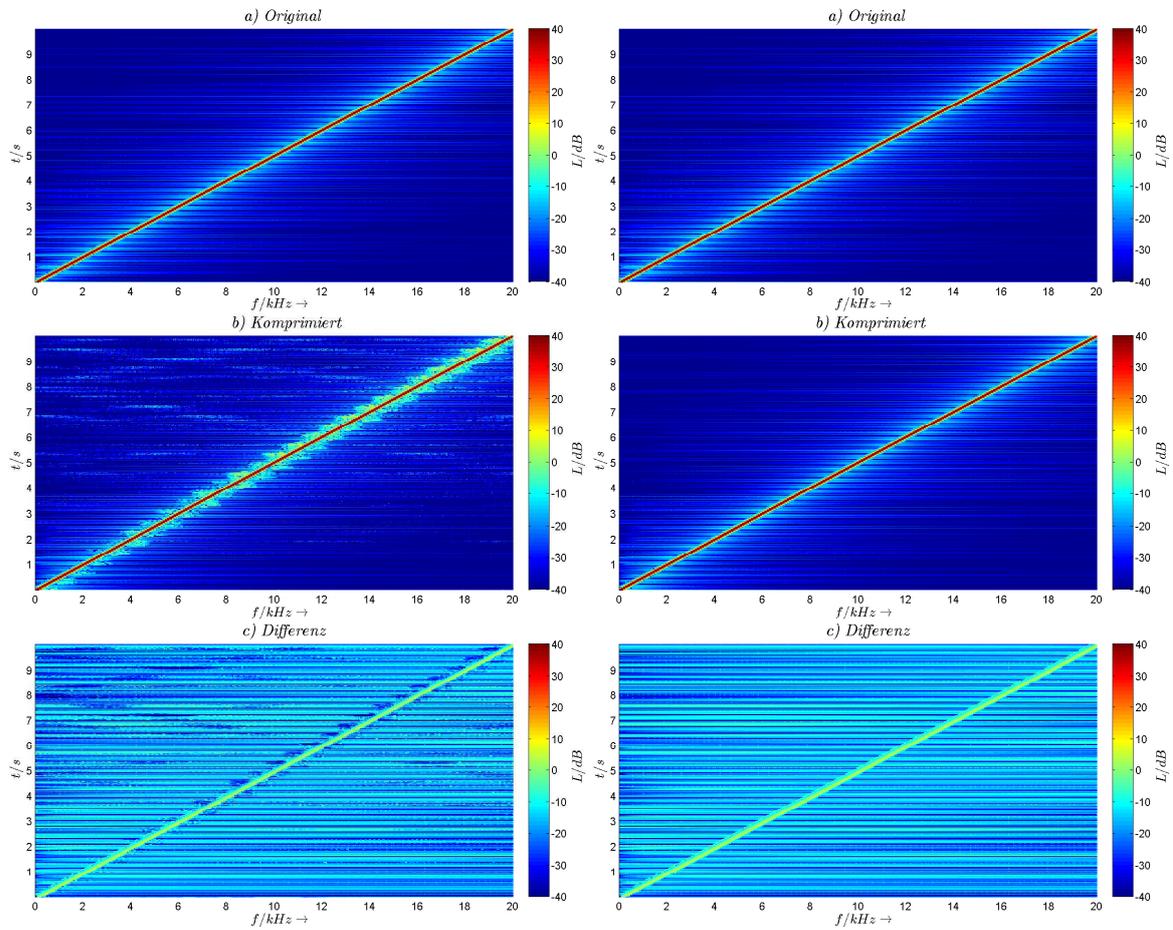


Abbildung 4.22.: Spektrogramme \mathcal{S}^{Swe} mit $v_B = 64$ kbps (Simulation)

Abbildung 4.23.: Spektrogramme \mathcal{S}^{Swe} mit $v_B = 384$ kbps (Simulation)

Die Abbildung 4.25 und Abbildung 4.24 zeigen die Ergebnisse für das Musikstück \mathcal{S}^{Bra} . Obwohl sich Musik wie ein Tiefpasssignal verhält (stärkere Signalleistung bei niedrigeren Frequenzen), erstrecken sich Frequenzanteile von \mathcal{S}^{Bra} bis 20 kHz. Bei dem originalen und dem komprimierten Spektrogramm sind bei $v_B = 384$ kbps mit bloßem Auge keine Unterschiede zu erkennen. Nur die Differenz zeigt eine scheinbar zufällige Reduktion der Signalanteile. Bei $v_B = 64$ kbps sind die hohen Frequenzen deutlich vermindert, was aus der schwachen Signalleistung und dem Mangel an Bits pro Rahmen resultiert.

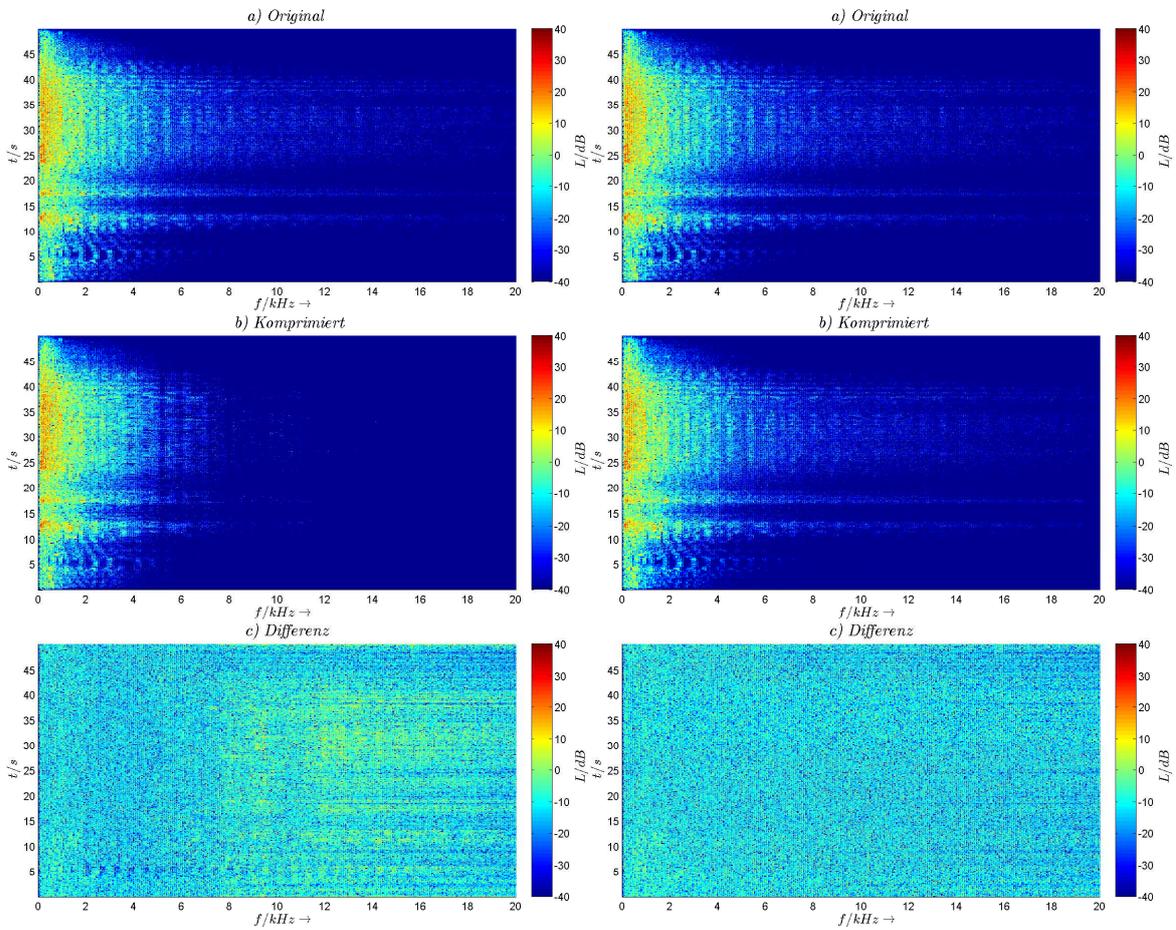


Abbildung 4.24.: Spektrogramme \mathcal{S}^{Bra} mit $v_B = 64$ kbps (Simulation)

Abbildung 4.25.: Spektrogramme \mathcal{S}^{Bra} mit $v_B = 384$ kbps (Simulation)

Den Abbildungen 4.25 und 4.24 sind die Ergebnisse für das Musikstück \mathcal{S}^{Joh} zu entnehmen. Hier handelt es sich um ein Alternative-Rock Lied mit starkem Einfluss einer elektrischen Gitarre. Diese erzeugt deutlich mehr Spektralanteile bei höheren Frequenzen als das vorige Musikstück. Dementsprechend werden diese bei $v_B = 64$ kbps so gut wie gar nicht mehr codiert. Hier sollten dem Hörer die größten Unterschiede auffallen.

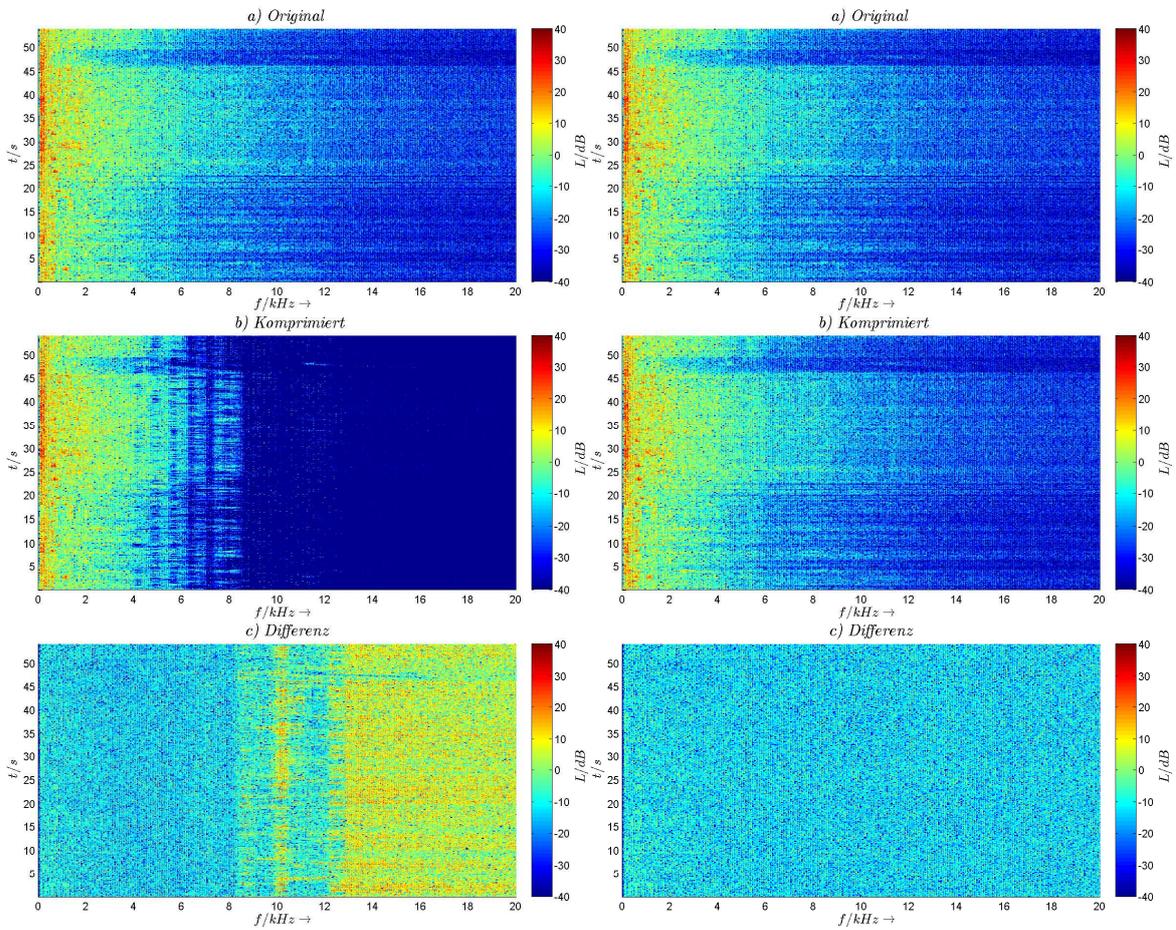


Abbildung 4.26.: Spektrogramme \mathcal{S}^{Joh} mit $v_B = 64$ kbps (Simulation)

Abbildung 4.27.: Spektrogramme \mathcal{S}^{Joh} mit $v_B = 384$ kbps (Simulation)

Das letzte Musikstück ist aus dem Bereich der Elektromusik. Anhand der original Spektrogrammen aus Abbildung 4.29 und Abbildung 4.28 ist zu erkennen, dass dieses weniger Bandbreite einnimmt. Hohe Frequenzen kommen so gut wie gar nicht vor. Lediglich in den ersten 30 Sekunden sind einige breitbandigere Signale wahrzunehmen. Sie entstammen aus einem Sprechgesang mit verstärkten Höhen. Diese werden bei $v_B = 64$ kbps weniger codiert woraus stärkere Verluste resultieren.

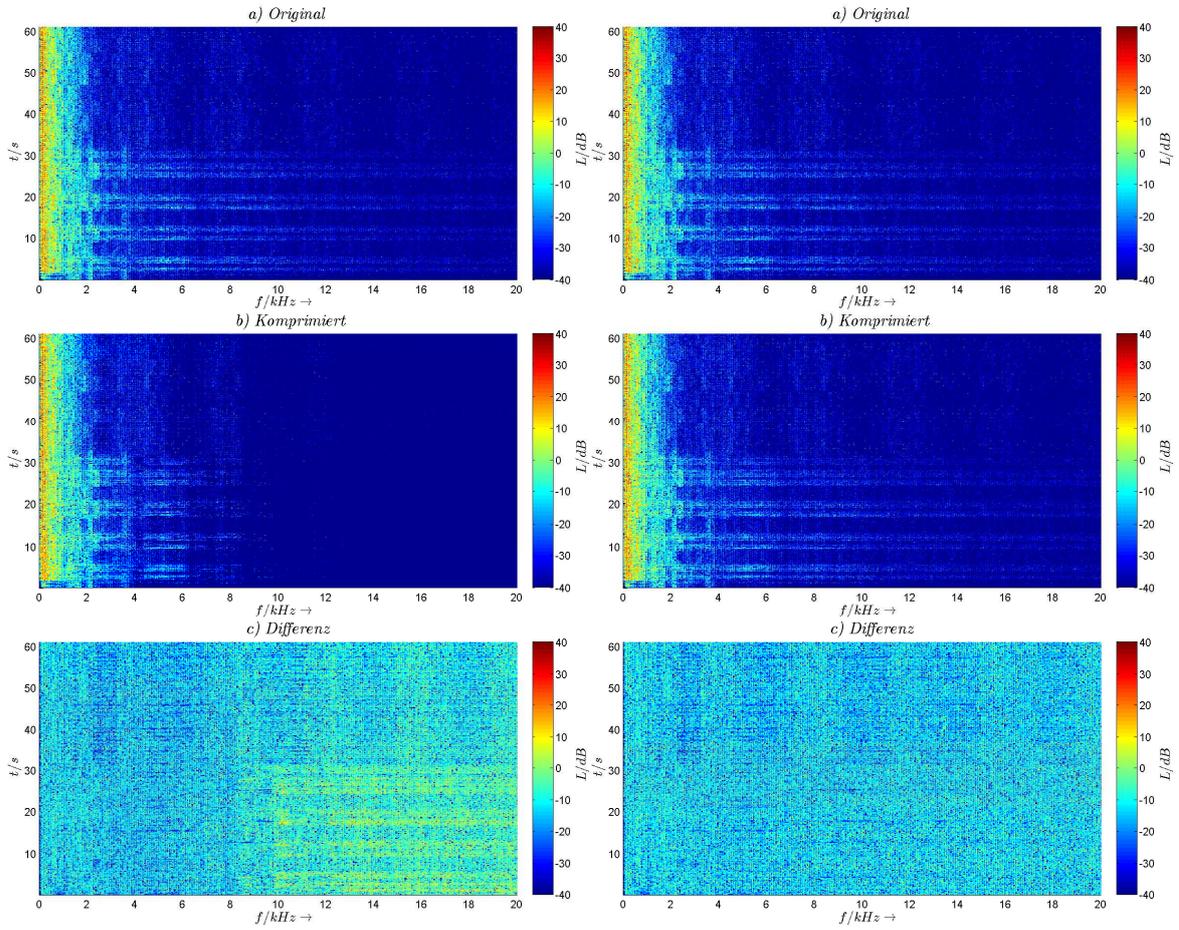


Abbildung 4.28.: Spektrogramme \mathcal{S}^{Kal} mit $v_B = 64$ kbps (Simulation)

Abbildung 4.29.: Spektrogramme \mathcal{S}^{Kal} mit $v_B = 384$ kbps (Simulation)

5. Implementierung

Dieses Kapitel zeigt die hardwarenahe Implementierung des in Kapitel 4 (MPEG-1 Audio Codierung) entwickelten und simulierten MPEG-1 Verfahrens. Dabei kommen zwei Boards zum Einsatz, eins als Encoder und eins als Decoder. Die Möglichkeit das gesamte System auf nur einem Board durchzuführen wird ebenso geprüft. Hierbei muss jedoch damit gerechnet werden, dass die Echtzeit-Bedingung nicht eingehalten werden kann, da der Rechenaufwand mit der Bitrate steigt. Da Hardware der Firma *Texas Instruments* verwendet wird, kommt als Entwicklungsumgebung *Code Composer Studio* Version 5.4 zum Einsatz. Diese bietet entsprechende Hardware-Programmoptimierungen durch dessen Compiler an. Bei den implementierten Berechnungen wurde stets die höchste Optimierungsstufe 3 (-O3) genutzt.

5.1. Verwendete Hardware

Die Implementierung wird auf dem Entwicklungsboard TMS320C6713 DSP Starter Kit (DSK) durchgeführt, welches in Abbildung 5.1 dargestellt ist. Auf diesem Board ist der DSP TMS320C6713B von *Texas Instruments* vorhanden. Dieser ist ein Floating-Point Prozessor mit acht parallelen Recheneinheiten und drei Pipeline-Stages. Bei einer Taktfrequenz von 225 MHz erreicht der C6713 bis zu 1350 Mega-Floating Point Operationen pro Sekunde (FLOPS). Damit bewegt sich dieser im unteren Mittelfeld. Zum Vergleich besitzt beispielsweise der aktuellere C6678 Prozessor bereits 160 Giga-FLOPS bei 1,25 GHz¹².

¹²Stand: Juni 2015, Quelle: www.ti.com

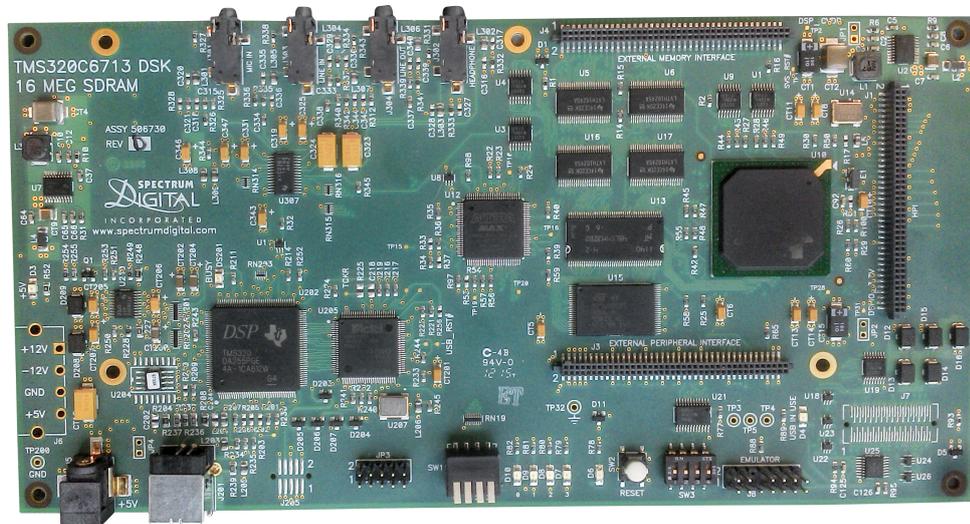


Abbildung 5.1.: Entwicklungsboard C6713 DSK der Firma *Spectrum Digital*

Abbildung 5.2 zeigt das Blockschaltbild des DSPs. Von besonderer Bedeutung ist der Multichannel Buffered Serial Port (McBSP) und der Enhanced Direct Memory Access (EDMA) Controller. Eine Schnittstelle zwischen digitaler und analoger Audiosignale liefert das DSK-Board mit dem on-board AIC23-Codec. Dieser Audio-Codec beinhaltet Sigma-Delta Wandler, durch eine Initialisierung abgeglichenen Offset-Verstärker und digital realisierte Anti-Aliasing- sowie Interpolations-Filter. Die digitalisierten Audiodaten werden vom AIC23-Codec mittels Serial Peripheral Interface (SPI) an den DSP internen McBSP übertragen. Durch das External Memory Interface (EMIF) des DSPs ist eine 32 Bit-breite Kommunikation mit der auf dem Board vorhandenen Peripherie möglich. Der on-board CPLD¹³ dient als externe Glue-Logic¹⁴. Über diesen wird ein Multiplexer angesteuert, welcher die beiden McBSPs entweder auf den AIC23-Codec oder auf die Pins der Steckerleiste J3 für externe Kommunikation des Boards routet.

¹³Ein CPLD (Complex Programmable Logic Device) ist ein programmierbarer Logik-Baustein.

¹⁴Eine Glue Logic dient zur Verbindung inkompatibler Hardwarebausteine.

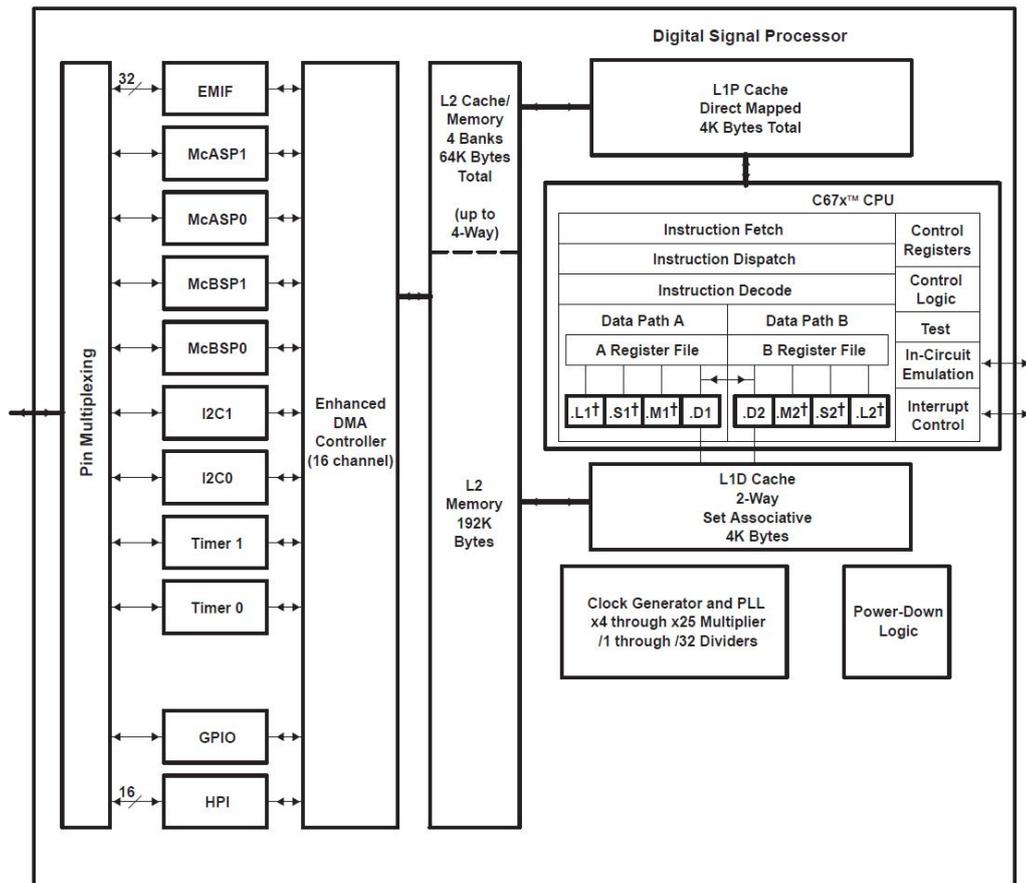
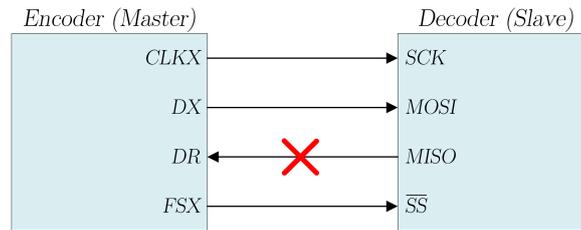
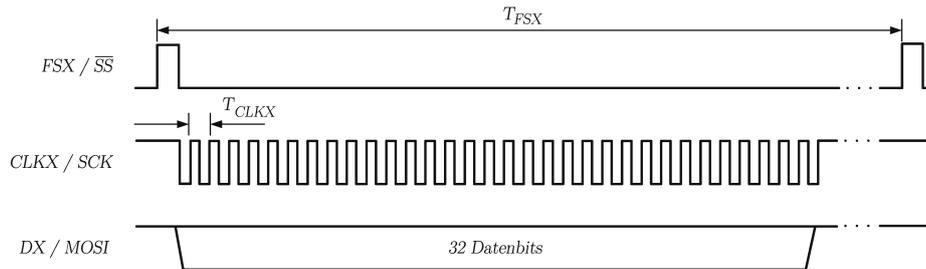


Abbildung 5.2.: Blockschaltbild des TMS320C6713B [Tex06c]

5.2. Datenübertragung

Für die Übertragung des komprimierten Audiostreams zwischen den beiden Boards wird das SPI-Protokoll verwendet. Dies ist ein vier-Leiter-Bus mit Master-Slave Prinzip, welcher durch den McBSP realisiert wird. Abbildung 5.3 zeigt die vier Signalleitungen zwischen Master und Slave sowie die zeitliche Darstellung der Signale. Es wurde der Clock-Mode 3 der vier möglichen SPI-Modes gewählt. Dieser Mode bedeutet, dass die gesendeten Datenbits DX und das Taktsignal $CLKX$ während einer inaktiven Zeit einen High-Pegel besitzen. Bei aktiver Phase übermittelt der McBSP Daten mit steigender Flanke des Taktsignals und empfängt Daten bei fallender Flanke. Das Empfangen der Daten wird nicht verwendet. Eine aktive Phase wird mit dem Frame-Select Signal FSX eingeleitet und bleibt in diesem Fall für eine Dauer von 32 Taktsignalen aktiv.

(a) Blockschaltbild – Signal *MISO/DR* wird nicht verwendet

(b) Zeitliche Darstellung der Bussignale

Abbildung 5.3.: Darstellung des SPI-Bus

Die Frequenz des Signals *CLKX* ergibt sich zu

$$F_{CLKX} = \frac{F_{CPU}/2}{CLKGDV + 1}, \quad (5.1)$$

wobei $CLKGDV = 0 \dots 255$ der 8-Bit Taktteiler des McBSP-Sample Rate Generator Registers (SRGR) repräsentiert. Die Taktperiode des Signals *FSX* ergibt sich aus ganzzahlig-vielfachem von $T_{CLKX} = 1/F_{CLKX}$ zu

$$T_{FSX} = (FPER + 1) \cdot T_{CLKX}, \quad (5.2)$$

mit dem 12 Bit Faktor $FPER = 0 \dots 4095$ des SRG Registers. Mit den durch die einstellbaren Werte möglichen Taktraten/-Frequenzen entsteht ein Problem, welches folgend erläutert wird.

Der Encoder DSP bekommt Abtastwerte vom AIC23-Codec mit der von einem Quarz generierten Frequenz von 48 kHz. Für die Verarbeitung werden 384 Abtastwerte gepuffert, woraus eine Updaterate von

$$T_{Update} = \frac{384}{48 \text{ kHz}} = 8 \text{ ms} = T_{EZ} \quad (5.3)$$

resultiert. Diese Updaterate kennzeichnet ebenso die Grenze für die Echtzeitbedingung T_{EZ} . Das Senden des Encoders muss bei jeder Bitrate bzw. jedem Kompressionsgrad 8 ms betragen. Abbildung 5.4 verdeutlicht dies.

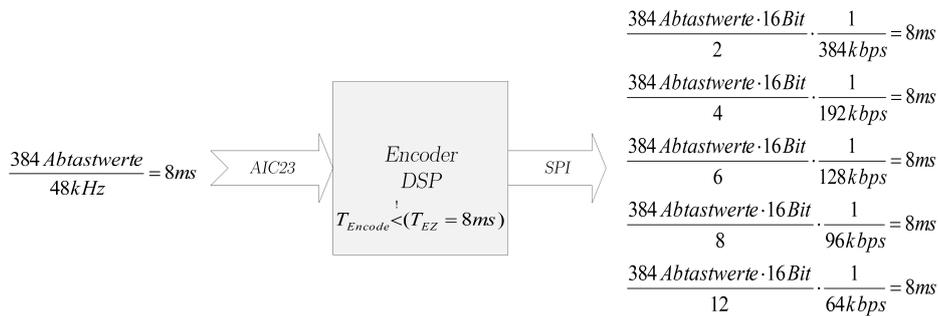


Abbildung 5.4.: Darstellung der Blockdauer von 8 ms bei unterschiedlichen Bitraten/Kompressionsfaktoren

Die Übertragungsraten, die der MPEG-1 Encoder realisieren soll, sind mit dem gegebenen Systemtakt von $F_{CPU} = 225 \text{ MHz}$ und den Gleichungen 5.1 und 5.2 jedoch nicht realisierbar. Es würde zu Asynchronität zwischen den empfangenen Abtastwerten und gesendeten Bits kommen. Synchronität kann nur erreicht werden, wenn der Takt des AIC23-Codex für die SPI zwischen beiden DSPs verwendet wird. Abbildung 5.5 zeigt das Blockschaltbild beider DSK-Boards mit interner Beschaltung zwischen AIC23-Codex und der DSP-Peripherie.

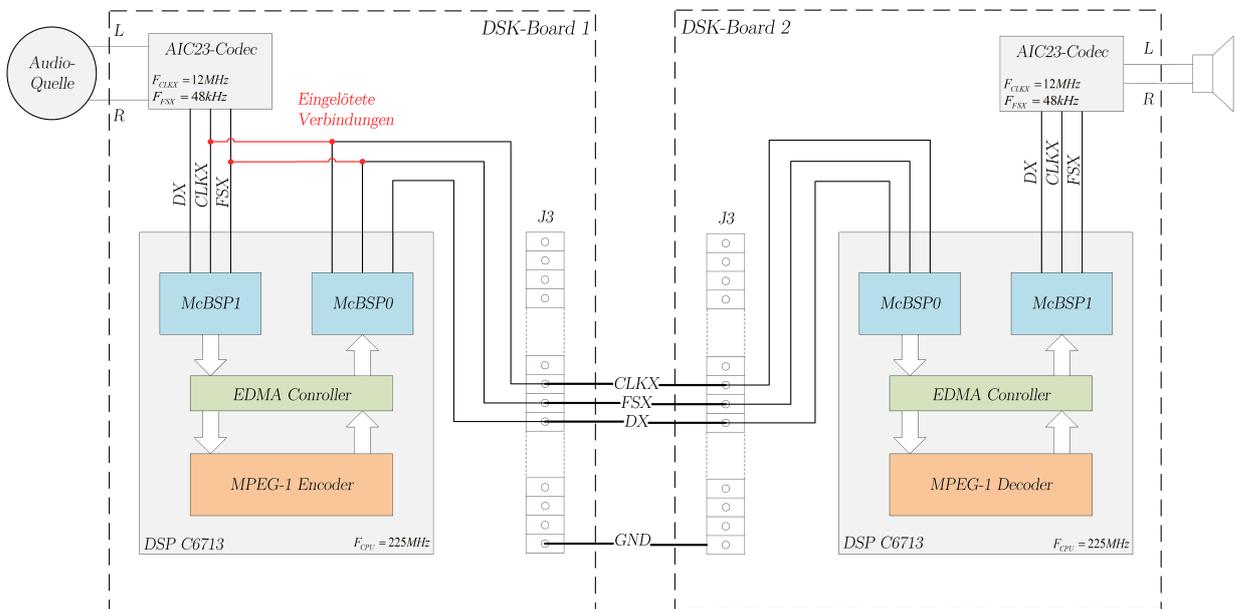


Abbildung 5.5.: Blockschaltbild der DSK-Boards und Beschaltung der AIC23-Codex mit dem DSP

Der AIC23-Codec des DSK-Board 1 sendet Abtastwerte im 32 Bit Rahmen mit $F_{CLKX} = 12$ MHz und $F_{FSX} = 1/T_{FSX} = 48$ kHz an den McBSP1. Dabei stehen die niederwertigeren 16 Bit für den linken und die höherwertigeren 16 Bit für den rechten Kanal. Durch das Bit $CLKSM$ des McBSP0-Sample Rate Generator Register (SRGR) und das Bit $FSXM$ des McBSP0-Pin Control Register (PCR) können die SPI-Signale $CLKX$ und FSX als Eingänge definiert werden. Diese werden mit den SPI-Signalen vom AIC23-Codec verbunden (rote Verbindungen in Abbildung 5.5). Der SPI-Slave, das DSK-Board 2, wird durch den McBSP0 realisiert. Der McBSP1 des Decoder-DSPs übermittelt die rekonstruierten Abtastwerte an den AIC23-Codec. Theoretisch beträgt die Übertragungsrate bei dieser Lösung statisch $v_B = 12$ Mbps. Die effektive Bitrate wird jedoch nur durch den Encoder-Prozess mit der dynamischen Bitzuweisung bestimmt. Alles darüber hinaus ist redundant. Zudem besteht der Schwerpunkt dieser Arbeit darin Audiosignale in Echtzeit zu komprimieren und dabei die Rechenarbeit auf einen Encoder und einem Decoder aufzuteilen.

Welche Funktionalität noch nicht erwähnt wurde, ist die Signalisierung, wann die Daten vom McBSP bereit stehen und verarbeitet werden können. Genauso muss eine Signalisierung erfolgen, wann die Daten vollständig durch den McBSP gesendet wurden und aktualisiert werden müssen. Hierfür existieren zwei Methoden. Eine Möglichkeit ist es, einen Interrupt vom CPU erzeugen zu lassen. Dies hat den Nachteil, dass die Bearbeitung des Hauptprogramms bei jedem Datentransfer unterbrochen wird. Die zweite Möglichkeit, welche realisiert wird, bietet die Verwendung des EDMA-Controllers. Dieser speichert die abgehenden und ankommenden Daten der McBSPs vollständig nebenläufig im Speicher zwischen. Die CPU erhält vom EDMA-Controller dann einen Interrupt, wenn ein ganzer Block an Daten vollständig transferiert und im Speicher zur Verfügung steht. Damit wird wichtige Rechenzeit gespart. Abbildung 5.6 zeigt das Prinzip-Flussdiagramm des Encoders unter Verwendung des EDMA-Controllers. Dabei beträgt der Datentransfer für beide McBSPs Blockweise 384 Werte, zu je 32 Bit.

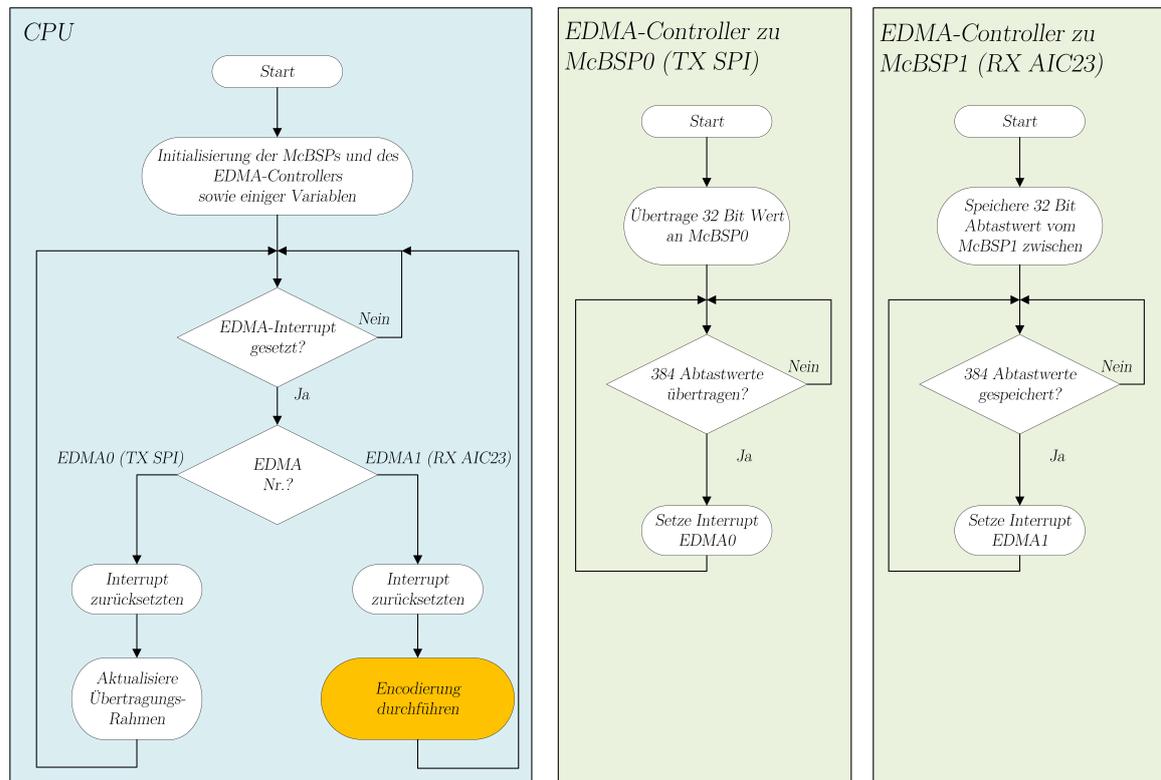


Abbildung 5.6.: Flussdiagramm des Encoders unter Verwendung des EDMA-Controllers. Der Encoder-Vorgang (farblich gekennzeichnet) wird im Abschnitt 5.4 detailliert dargestellt.

Abbildung 5.7 und Abbildung 5.8 zeigen die gemessenen SPI-Bussignale zwischen den DSK-Boards. Wie erläutert, stammen die Signale *FSX* (CH1) und *CLKX* (CH2) vom AIC23-Codec.

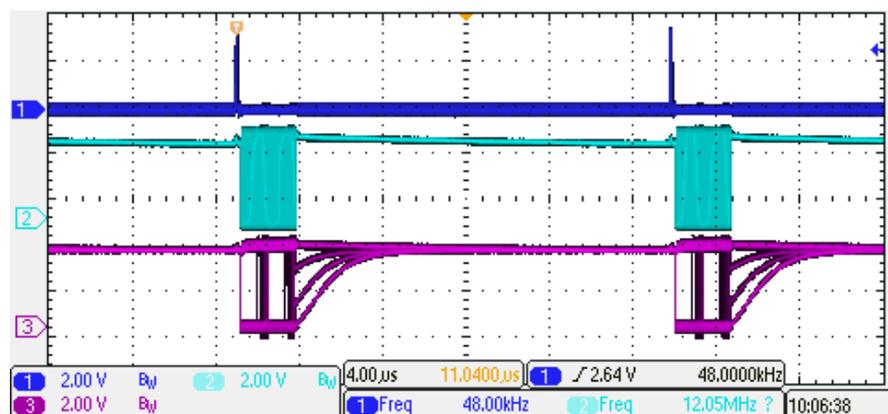


Abbildung 5.7.: Messung des SPI Bus (Messdauer von zwei *FSX* Signalen)
CH1: *FSX*, CH2: *CLKX*, CH3: *DX*

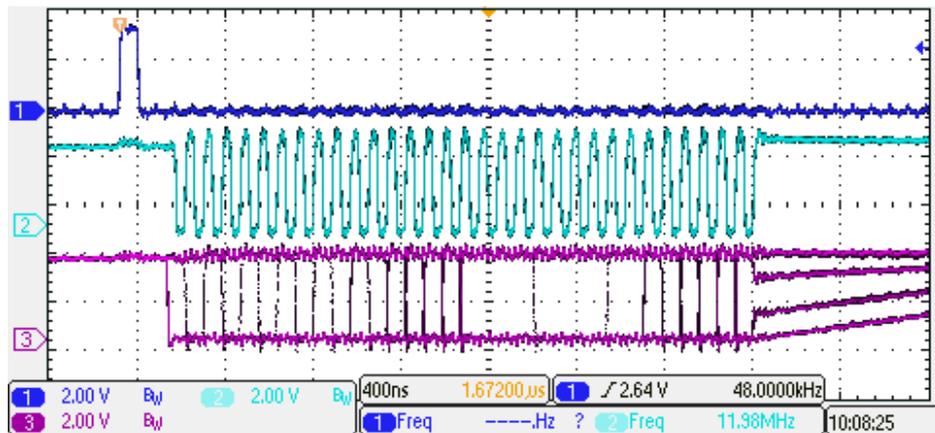


Abbildung 5.8.: Messung des SPI Bus (Messdauer eines *FSX* Signals)
 CH1: *FSX*, CH2: *CLKX*, CH3: *DX*

5.3. Filterbank

Da die Grundlagen der Filterbank (Abschnitt 3.6) sowie die Implementierungsdarstellung (Abbildung 4.5 b) und Abbildung 4.21 b)) bereits vorgestellt wurden, folgen in diesem Unterkapitel Messungen an der implementierten Filterbank. Die Realisierung in C ist dem Datenträger beigelegt (Anhang C). Im Folgenden wird der Amplitudengang der Analyse-Filterbank gemessen. Eine gängige Methode ist die Messung über den AIC23-Codec des DSK-Boards mit einem Audioanalysator. Die hohe Sperrdämpfung (≈ 110 dB vgl. Abbildung 4.6 a) aus Abschnitt 4.1) der Filterbank übersteigt jedoch die mögliche Auflösung der auf dem DSK-Board vorhandenen 16 Bit Wandler. Eine Alternative bietet die direkte Auswertung der Filterbankdaten durch den McBSP. Hierfür werden die McBSP-Signale statt zum AIC23-Codec auf die Steckerleiste J3 geroutet, auf welche ein Adapterboard gesteckt wird, vgl. Abbildung 5.9. Auf dem Audioanalysator ist eine Setup-Datei gegeben, die eine 32-Bit Messung über I²S¹⁵ durchführt. Die McBSPs sind ebenso für ein 32-Bit I²S-Protokoll konfiguriert.

¹⁵I²S steht für Inter-IC Sound und ist ein Voll duplex Bus-Protokoll zum Austausch von digitalen Daten zwischen zwei Geräten.

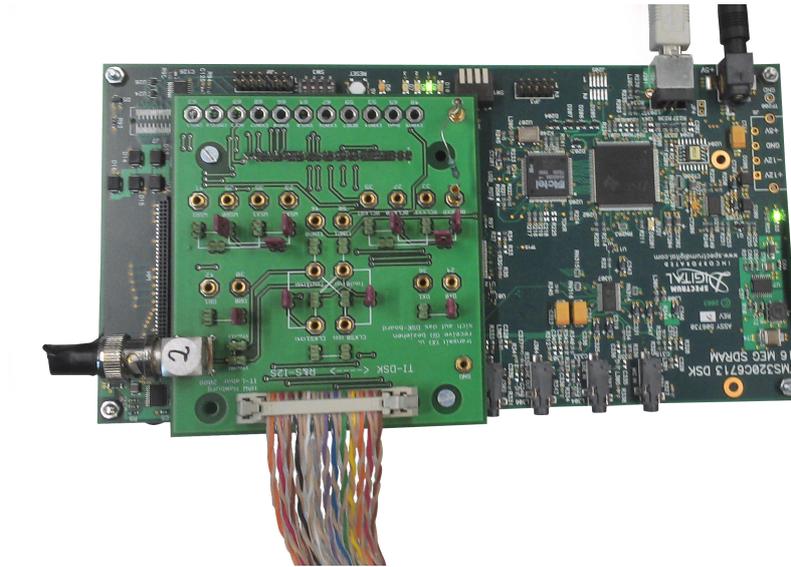


Abbildung 5.9.: DSK-Board mit aufgestecktem Adapterboard des IT-Labors an der HAW zur Messung eines Amplitudengangs mit 32 Bit Auflösung über I²S

Die gemessenen Amplitudengänge der unteren 16 Analyseteilbänder ist in Abbildung 5.10 dargestellt. Die Messung der oberen 16 Teilbänder zeigt Abbildung 5.11. Diese stimmen mit den simulierten Amplitudengängen in Abbildung 4.6 a) aus Abschnitt 4.1 gut überein. Auch die Sperrdämpfung von ungefähr 110 dB werden erreicht.

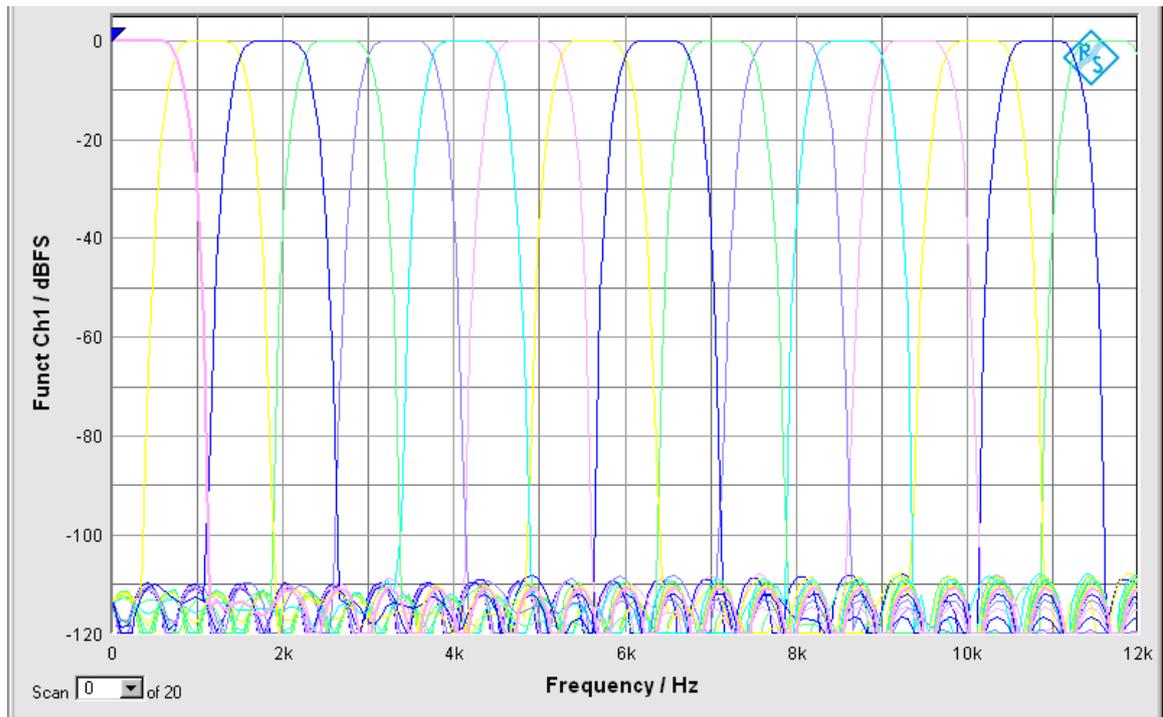


Abbildung 5.10.: Amplitudengänge der gemessenen unteren 16 Teilbänder

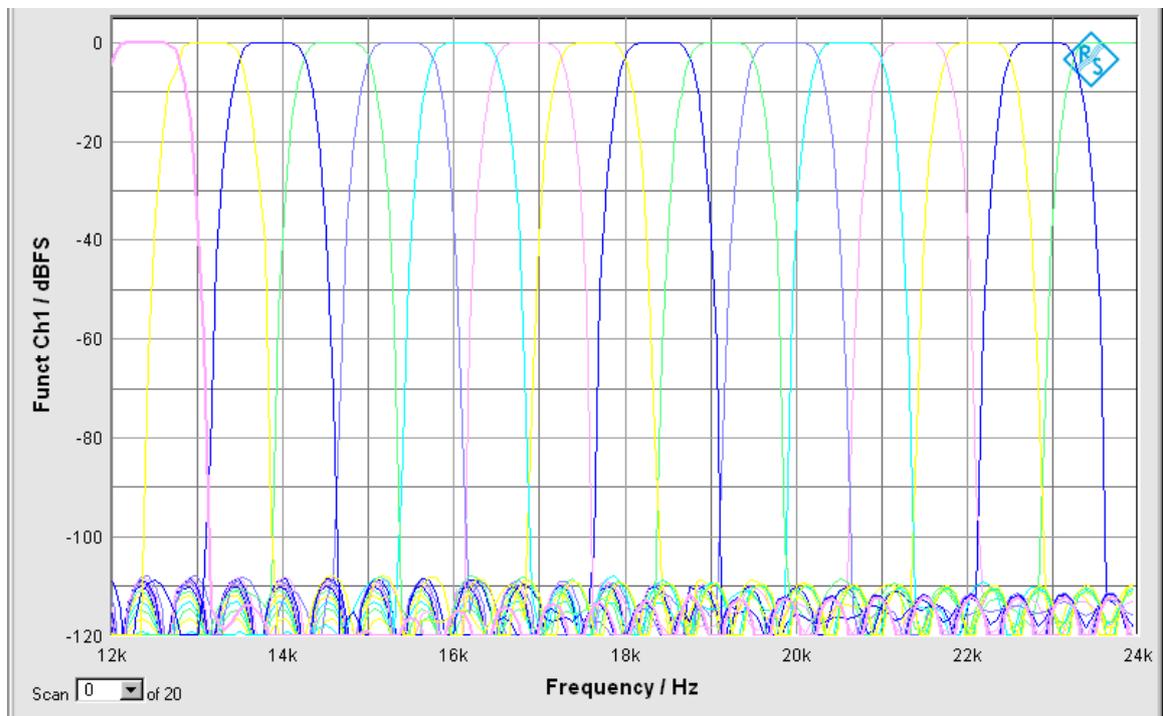


Abbildung 5.11.: Amplitudengänge der gemessenen oberen 16 Teilbänder

Als nächstes wird die Verzerrungsfunktion $F_{dist}(e^{j\omega})$, welche die gesamt-Übertragungsfunktion der Filterbank darstellt, gemessen. Hierfür wird der Analyse-Filterbank die Synthese-Filterbank nachgeschaltet. Diese Messung wird über den AIC23-Codec durchgeführt und ist ebenso dem Datenträger beigefügt. Das Ergebnis ist in Abbildung 5.12 dargestellt. Ein Vergleich mit der simulierten Verzerrungsfunktion in Abbildung 4.6 b) zeigt kleine Unterschiede. Neben einer Welligkeit ist ein Abknicken der gemessenen Verzerrungsfunktion bei 20 kHz erkennbar. Dies resultiert aus dem Übertragungsverhalten des DSK-Boards, welches durch das Anti-Aliasing- und Interpolations-Filter des AIC23-Codecs mit dem der Filterbank überlagert. Mit einem Ausschlag von 0,086 dB spielt dies jedoch keine signifikante Rolle.

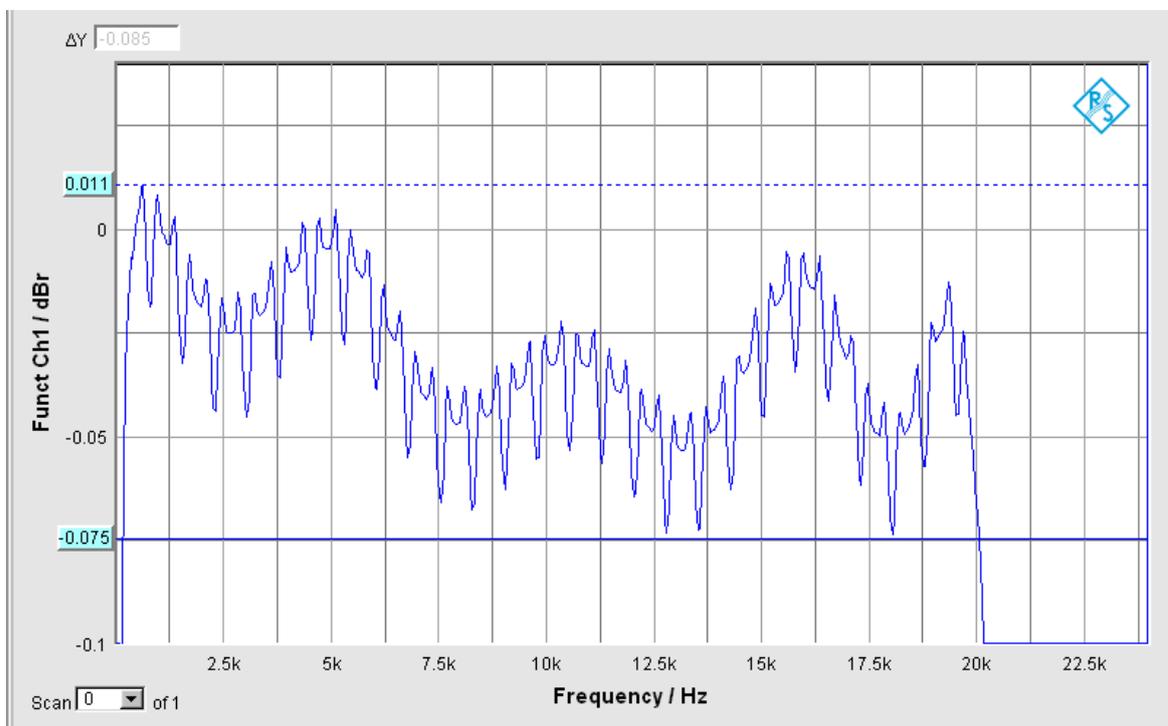


Abbildung 5.12.: Gemessene Verzerrungsfunktion der Filterbank

5.4. Audio Encoder

Der Encoder und dessen Methodik wurden in Kapitel 4 erläutert. In diesem Unterkapitel wird die auf der Simulation basierende Realisierung auf dem DSP vorgestellt. Das entwickelte CCS-Projekt ist dem Datenträger im Anhang beigelegt. Die implementierte Abfolge der Encoder-Teilaufgaben zeigt Abbildung 5.13. Die Analyse-Filterbank wurde bereits im vorigen Abschnitt gezeigt. Das Betragsspektrum basiert auf dem Radix 2 Algorithmus. Dies ist ein mathematisches Lösungsverfahren für die FFT. Die 2 kennzeichnet dabei, dass die Anzahl der Eingangswerte einer Potenz der Basis 2 entsprechen, welches mit einer Länge von $N_{FFT}=512$ erreicht wird. Ein in Assembler-optimierter Algorithmus für eine komplexe Floating-Point Radix 2 FFT ist durch Texas Instruments entwickelt worden. Diese ist in [CR10, S. 295 ff] beschrieben und wird in dieser Arbeit verwendet. Den Betrag des komplexwertigen Spektrums erhält man durch Gleichung 4.9. Eine Messung für $N_{FFT}/2=256$ Betragsbildungen hat eine Rechendauer von $T_{Abs-C}=319 \mu s$ ergeben. Diese konventionelle Methode kann für eine Einhaltung der Echtzeitbedingung kritisch sein. In [Int15] ist eine Methode gegeben, die deutlich rechen effizienter arbeitet. Dabei wird der Betrag durch

$$|X(k)| \approx \alpha \cdot \max(|\operatorname{Re}\{\underline{X}(k)\}|, |\operatorname{Im}\{\underline{X}(k)\}|) + \beta \cdot \min(|\operatorname{Re}\{\underline{X}(k)\}|, |\operatorname{Im}\{\underline{X}(k)\}|) \quad (5.4)$$

approximiert. Dabei gilt für die Konstanten $\alpha = 0,960433870103$ und $\beta = 0,397824734759$. Diese Werte entstammen aus der Tabelle in [Int15] und ergeben für die Approximation die kleinste Fehlerleistung. Mit $T_{Abs-A}=5,7 \mu s$ wurde somit fast 1/60 der Rechenzeit eingespart.



Abbildung 5.13.: Flussdiagramm des Encoders

Der nächste Schritt ist die Berechnung des Leistungspegels und die Normierung auf 96 dB. Da im Folgenden auch mit entlogarithmierten Leistungspegeln gerechnet wird,

wird der Betrag auf den entlogarithmierten Wert von 96 dB normiert

$$|X(k)| = \frac{|X(k)|}{X_{max}} \cdot 10^{\frac{96 \text{ dB}}{20}}. \quad (5.5)$$

Den Leistungspegel des Betrags erhält man mit Gleichung 4.10. Es zeigte sich, dass eine Berechnung von 256 Logarithmen $T_{Log-C}=1,66$ ms dauert. Dies ist ebenso kritisch für die Echtzeit. In [Int15] ist auch für Logarithmierungen eine rechen effizientere Lösung gegeben. Hierfür bedarf es an der Kenntnis über die DSP-Darstellungsform einer Zahl im 32-Bit Floating-Point Format mit Single-Präzision. Dabei setzt sich die Zahl X wie folgt zusammen [Tex06d]:

$$X = (-1)^{\text{Vorzeichen}} \cdot 1, \text{Mantisse} \cdot 2^{\text{Exponent}-127} \quad (5.6)$$

Abbildung 5.14 zeigt die 32-Bit Anordnung des Vorzeichen, des Exponents und der Mantisse.



Abbildung 5.14.: Darstellung der 32-Bit Single-Präzision, [Tex06d, S. 71]

s=Vorzeichen, e=Exponent, f=Mantisse

Die Logarithmierung wird zunächst zur Basis 2 durchgeführt. Dabei spielt das Vorzeichen durch die Betragbildung keine Rolle mehr.

$$\log_2(X) = \log_2(\text{Mantisse}) + (\text{Exponent} - 127) \cdot \log_2(2) \quad (5.7)$$

Geht man von einem möglichen Wertebereich $1 \leq \text{Mantisse} < 2$ aus, kann man den Logarithmus dieser Werte durch eine LUT realisieren. Je größer die Tabelle ausfällt, desto genauer ist das Ergebnis. Den Leistungspegel erhält man schließlich durch

$$\begin{aligned} 20 \cdot \log_{10}(X) &= 20 \cdot \frac{\log_{10}(2)}{\log_2(2)} \cdot [\text{LUT}[\text{Mantisse}] + (\text{Exponent} - 127) \cdot \log_2(2)] \\ 20 \cdot \log_{10}(X) &= C \cdot [\text{LUT}[\text{Mantisse}] + (\text{Exponent} - 127) \cdot \log_2(2)], \end{aligned} \quad (5.8)$$

wobei $C = 8,685889638$ und $\log_2(2)$ als Konstanten implementiert werden. Die realisierte C-Routine ist in Listing 5.1 dargestellt. Dabei wurde eine LUT mit 1000 Werten verwendet. Mit dieser Methode dauert die Berechnung nur noch $T_{Log-A}=14,32 \mu\text{s}$. Dies entspricht weniger als dem 1/120 der Rechenzeit gegenüber der konventionellen Methode.

```

breaklines
1 #include "init.h"           // variables initialisation
2
3 void calc_mag_trick(void){
4     /* local variables */
5     short i_for=0;
6     float mantissa_X, log_mantissa_X;
7     short exponent_X;
8     int INT_X;
9
10    for (i_for=0;i_for<NFFT/2;i_for++){
11        INT_X=(*(int*)&X[i_for]);           // cast to int
12        exponent_X=((INT_X & 0x7f800000)>>23); // get exponent
13        mantissa_X=(INT_X & 0x007fffff)/8388608.0; // get mantissa
14        log_mantissa_X=LuT[(short)(((mantissa_X*1000)))] // log of mantissa
15        /* now calc. power level of X in dB */
16        X_dB[i_for]=((log_mantissa_X+(float)((exponent_X-127)
17            *LOG2VALUE))*CONSTANT_FACTOR_LOG_TO_20LOG10);
18    }
19 }

```

Listing 5.1: C-Routine „calc_mag_trick.c“ zur Bestimmung des Leistungspegels

Die Berechnung der Skalenfaktoren, der Schalldruckpegel sowie die Ermittlung der Maskierer und der individuellen Mithörschwellen stellen keine zeitkritische Umsetzung in der Hardware dar und können nach der simulierten Methode realisiert werden.

Die Implementierung der globalen Mithörschwelle nach Gleichung 4.24 führt jedoch zu einer immensen Rechenzeit von $T_{g-C}=44$ ms. Dies liegt daran, dass alle Mithörschwellen sowie die absolute Hörschwelle entlogarithmiert aufsummiert werden müssen. Dabei kommen durch die Anzahl der Werte aus Tabelle 4.4 und den Frequenzgruppen für tonale und nicht-tonale Maskierer insgesamt $102 \cdot 2 \cdot 25=5100$ Entlogarithmierungen und Aufsummierungen zusammen. Zusätzlich kommen noch 102 Rückführungen in die dB-Skala. Auch eine Optimierung der Logarithmierungen führte zu keiner befriedigenden Rechendauer.

Abweichend vom MPEG-1 Standard wird daher die globale Mithörschwelle durch den jeweils größten Wert aus absoluter Hörschwelle, tonaler und nicht-tonaler Mithörschwelle gebildet. Diese Methode benötigt mit $T_{g-A}=137 \mu\text{s}$ nur etwa $1/320$ der Rechenzeit. Abbildung 5.15 a) zeigt die beiden unterschiedlich ermittelten globalen Mithörschwellen anhand des in Kapitel 4 dargestellten Simulationsbeispiels. In Abbildung 5.15 b) ist die Differenz beider Kurven gegeben. Die Bildung der globalen Mithörschwelle über die Maxima-Werte ist dabei kleiner gleich der Bildung durch entlogarithmierte Überlagerung. Die Spitzenwerte, die aus tonalen Maskierern resultieren, sind nahezu identisch und die Differenz geht gegen null. Dadurch, dass die Differenzen sich auf wenige dB belaufen, wird mit keinem signifikanten Einfluss auf die Bitzuweisung und die damit einhergehende Klangqualität gerechnet.

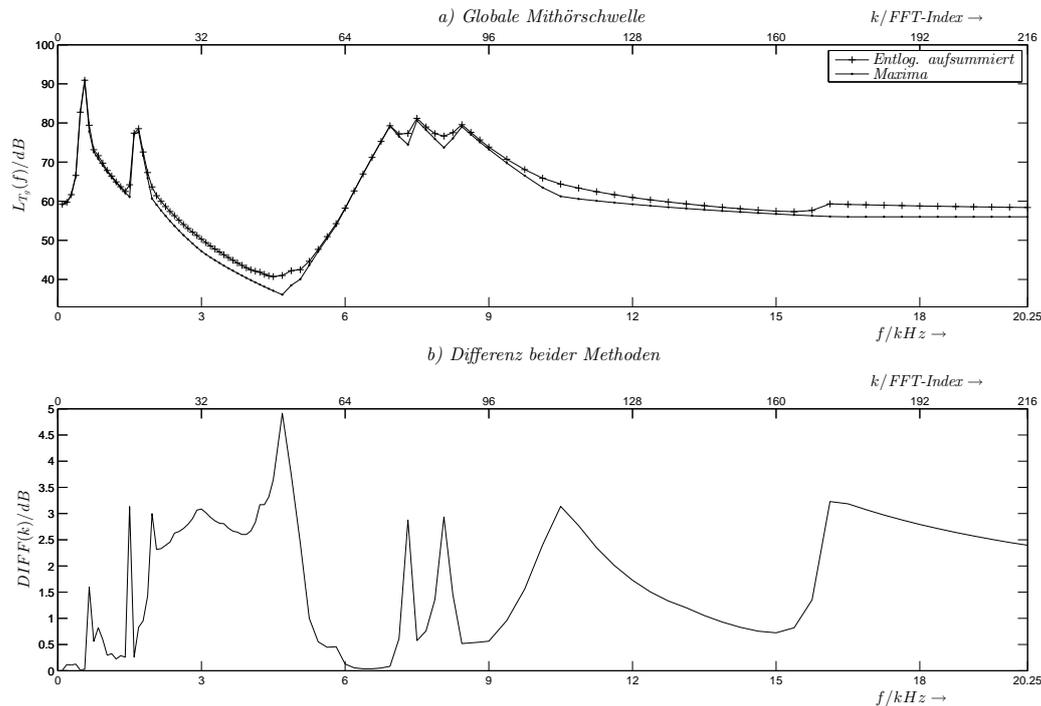


Abbildung 5.15.: Differenzen unterschiedlicher Berechnungen der globalen Mithörschwelle $L_{T_g}(f)$

Alle weiteren Berechnungen konnten ohne Abweichungen zur Simulation implementiert werden. Der im Laufe des Kapitel 4 gezeigte Vorgang zur Ermittlung der Mithörschwellen und der Bitanzahl je Teilband anhand des Test-Signal aus Gleichung 4.23, soll nun auf dem DSP durchgeführt und verglichen werden. Hierfür bietet CCS ein Tool, um verwendete Variablen grafisch darstellen zu können. Abbildung 5.16 bis Abbildung 5.23 zeigen den Vorgang vom Spektrum des Test-Signals bis zur Bitzuweisung bzw. den erreichten Mithörschwellen-Rauschabstand.

Das Spektrum $X(k)$ zeigt die drei überlagerten Sinussignale. Dabei ist der SNR geringer als bei der Simulation, vgl. Abbildung 4.11. Dies liegt an der geringeren Auflösung (16 Bit) des Eingangssignals gegenüber der Simulation. Der Schalldruckpegel $L_S(n)$ basiert auf dem Spektrum $X(k)$. Abgesehen vom geringeren SNR ist dieser nahezu identisch mit der Simulation, vgl. Abbildung 4.15 a). Ein Vergleich der minimalen Mithörschwelle $L_{T_{min}}(n)$ zeigt kleine Unterschiede. Diese können mit der Abweichung in der Ermittlung der globalen Mithörschwelle begründet werden. Der resultierende Signalverlauf des Signal-Mithörschwellenabstands $SMR(n)$ weicht bei genauer Betrachtung ebenso minimal von der Simulation ab, vgl. Abbildung 4.15 b). Damit einher gehen abweichende Werte für $BSPL(n)$ vor und nach der Bitzuweisung, vgl. Abbildung 4.17 b). Tabelle 5.1 listet die zugewiesenen Bits je Teilband auf. Die Differenz zur Simulation (Tabelle 4.5) zeigt, dass auf dem DSP im Mittel ein Bit, bzw. für

12 Abtastwerte je Teilband, 12 Bits weniger zugewiesen werden. Dabei zeigen beide Ergebnisse eine üppigere Bitverteilung bei tiefer liegenden Teilbändern. Der daraus resultierende Mithörschwellen-Rauschabstand $MNR(n)$ bewegt sich bei der Simulation sowie auch bei der Messung auf dem DSP bei einem Pegel von etwa 28 dB mit einem Ausschlag von weniger als ± 4 dB.

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
BSPL(i)	11	8	8	3	3	4	4	4	3	6	7	0	0	2	2	2
BSCF(i)	6	6	6	6	6	6	6	6	6	6	6	0	0	6	6	6
Diff. zur Simulation	+1	+1	-1	-1	0	0	-1	-1	-1	0	0	-2	-2	0	0	0
i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
BSPL(i)	2	2	2	2	2	2	2	3	3	2	2	2	2	2	2	2
BSCF(i)	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
Diff. zur Simulation	0	0	0	0	0	0	0	+1	+1	+2	0	0	0	+2	0	0

Tabelle 5.1.: Zugewiesene Bits für die 32 Teilbänder mit Differenz zur Simulation

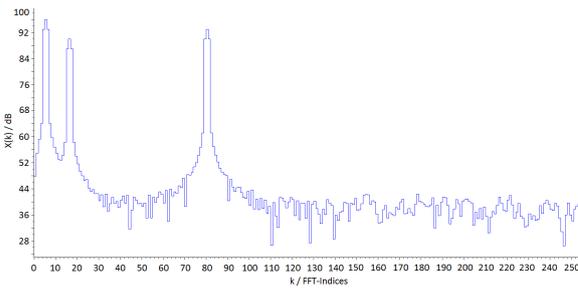


Abbildung 5.16.: $X(k)$

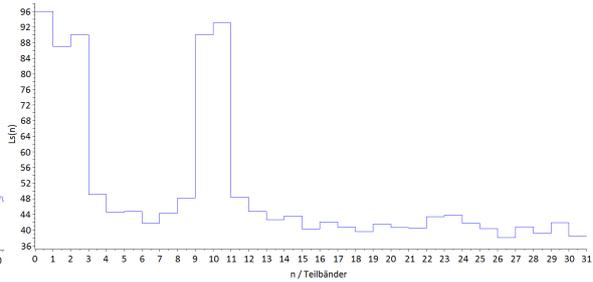


Abbildung 5.17.: $L_S(n)$

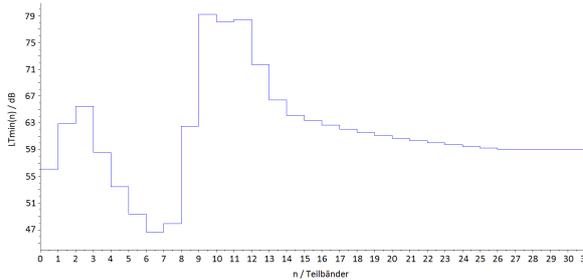


Abbildung 5.18.: $L_{T_{min}}(n)$

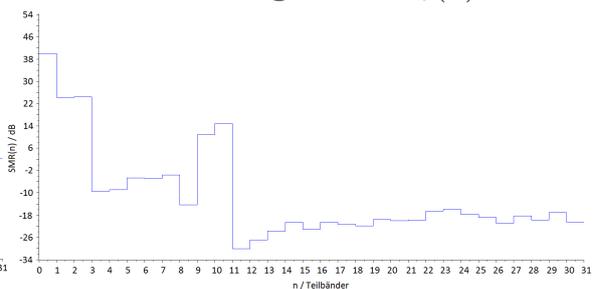


Abbildung 5.19.: $SMR(n)$

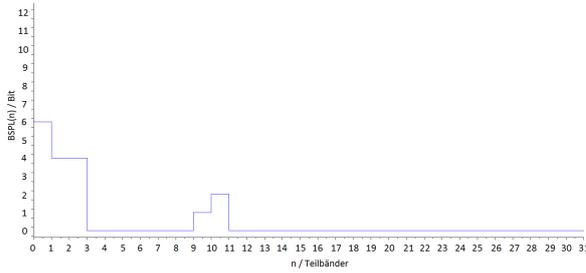


Abbildung 5.20.: $BSPL(n)$
vor der Bitzuweisung

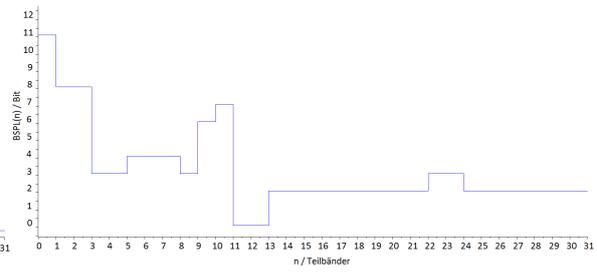


Abbildung 5.21.: $BSPL(n)$
für $v_B = 192$ kbps

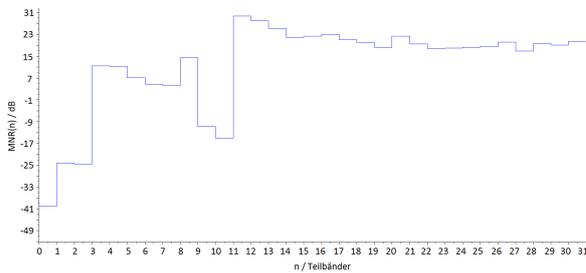


Abbildung 5.22.: $MNR(n)$
vor der Bitzuweisung

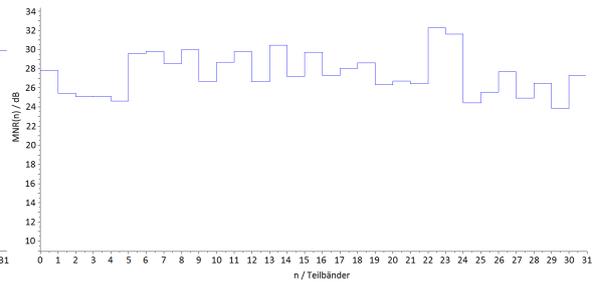


Abbildung 5.23.: $MNR(n)$
für $v_B = 192$ kbps

Die Rahmenbildung erfolgt wie in Kapitel 4 beschreiben. Zur Erkennung eines übertragenen Rahmens bedarf es einer Synchronisationsfolge. Während des Betriebs soll es möglich sein, zwischen originalen und komprimierten Bitstream umschalten zu können. Im unkomprimierten Fall werden lediglich die Abtastwerte übertragen. Um dies dem Decoder bekannt zu machen, wird eine vom MPEG-1 Standard abweichende Synchronisationsfolge verwendet. Für eine Startsequenz eines originalen Rahmens wird die zufällig gewählte Folge `0xC0C0AAAAAAAAAF0F0` verwendet. Für eine komprimierte Datenfolge wird die Präambel `0xAAAACCCCF0F0AAAA` vorangestellt.

5.5. Audio Decoder

Die Aufgabe des Decoders besteht darin, die in dem empfangenen Rahmen enthaltenen quantisierten Teilbandabstastwerte zusammenzufügen und mittels Synthese-Filterbank zu dem Audiosignal zu rekonstruieren. Das Flussdiagramm des Decoders ist in Abbildung 5.24 dargestellt. Der Datentransfer zwischen McBSPs und der CPU wird wie beim Encoder jeweils durch einen EDMA-Controller realisiert. Zunächst wird nach der Start-Sequenz im empfangenen Bitstream gesucht. Handelt es sich um die originalen Abtastwerte, werden diese direkt auf den EDMA-Buffer des McBSP1 geschrieben, vgl. hierzu auch Abbildung 5.5 mit dem Blockschaltbild beider DSK-Boards. Bei codierten Abtastwerten findet zunächst das Demultiplexen des empfangenen Rahmens zu 32 Teilbändern statt. Alle im Rahmen fehlenden Abtastwerte werden mit Nullen gefüllt, sodass eine Datenfolge wieder aus 32 Teilbändern, zu je 12 Abtastwerten, also insgesamt 384 Werten besteht. Im Anschluss folgt die Signal-Synthese und das beschreiben des EDMA-Buffers zur Übertragung an den AIC23-Codec.

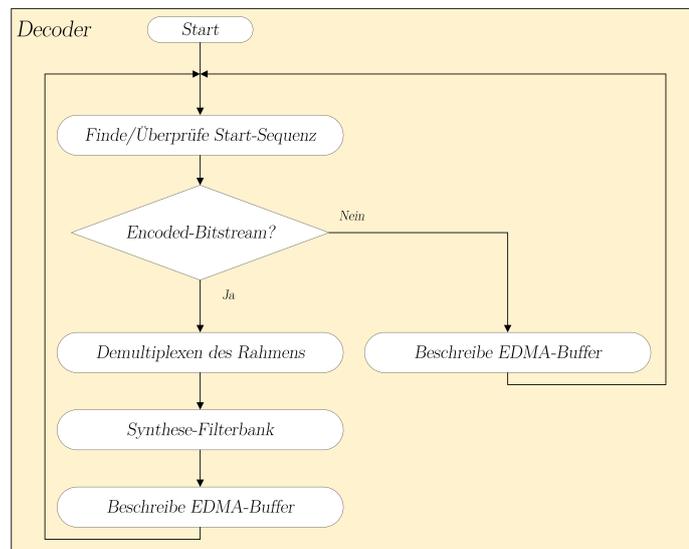


Abbildung 5.24.: Flussdiagramm des Decoders

Die Entwicklung einer DSP basierenden Bitratenreduktion von Audiosignalen auf der Grundlage des MPEG-1 Verfahrens ist nun abgeschlossen. Alle CCS-Projekte sind dem Datenträger zu entnehmen. Im Weiteren wird das System anhand von Messungen und Tests bewertet.

6. Messungen und Tests

Um ein datenreduziertes Audiosignal zu beurteilen, sind mehrere Verfahren möglich. In dieser Arbeit erfolgen neben der Darstellung von Spektrogrammen Messungen des Amplitudengangs. Ein abschließender Hörtest mit einigen Testpersonen dient zur subjektiven Beurteilung. Zuvor erfolgt eine Überprüfung der Echtzeit-Bedingung.

6.1. Echtzeit

Neben der separaten Entwicklung des Encoders und Decoders, wird auch die Realisierung auf nur einem Board überprüft. Dabei darf die Berechnungsdauer $T_{EZ}=8$ ms nicht überschritten werden. Die Messergebnisse für verschiedene Bitraten sind in Tabelle 6.1 dargestellt. Bis auf $v_B=384$ kbps ist eine echtzeitfähige Umsetzung auf nur einem DSK-Board möglich.

v_B /kbps	1 DSK-Board	2 DSK-Boards	
		Encoder	Decoder
64	6,1 ms	3,8 ms	$\approx 2,5$ ms
96	6,3 ms	3,9 ms	$\approx 2,5$ ms
128	6,7 ms	4,0 ms	$\approx 2,5$ ms
192	7,2 ms	4,2 ms	$\approx 2,5$ ms
384	$>8,0$ ms	4,6 ms	$\approx 2,5$ ms

Tabelle 6.1.: Messungen der totalen Berechnungsdauer für Echtzeitbedingung

6.2. Amplitudengänge

Die Darstellung der Amplitudengänge soll bitratenabhängige Verzerrungen aufzeigen. Die Messergebnisse sind in Abbildung 6.1 bis Abbildung 6.6 zu finden. Der Amplitudengang des DSK-Boards zeigt eine Welligkeit, die zuvor schon bei der Filterbank-Verzerrungsfunktion (Abbildung 5.12) auffiel und beschrieben wurde. Bei $v_B = 384$ kbps sind kaum Unterschiede auszumachen, lediglich zu hohen Frequenzen sind einige leichte Verzerrungen zu erkennen. Wahrnehmbare Verluste der Klangqualität sind für

ein durchschnittliches Gehör nicht zu erwarten. Nach weiterer Halbierung, $v_B = 192$ kbps, der Bitrate entstehen über dem gesamten Frequenzbereich leichte Verzerrungen. Der Ausschlag bleibt dabei noch unter 0,5 dB. Bei $v_B = 128$ kbps entstehen bereits Amplitudeneinbrüche bis zu 1,8 dB, wobei bei niedrigeren Frequenzen weniger Einbrüche auftreten. Wie zuvor bereits bei den Spektrogrammen beschrieben, ist dies der Einfluss der Frequenzgruppen nach Zwicker. Erste wahrnehmbare Klangeinbußen werden bei $v_B = 96$ kbps erwartet. Höhen werden zunehmend weniger quantisiert. Mit Einbrüchen von fast 7 dB sind die Auswirkungen bei $v_B = 64$ kbps noch stärker.

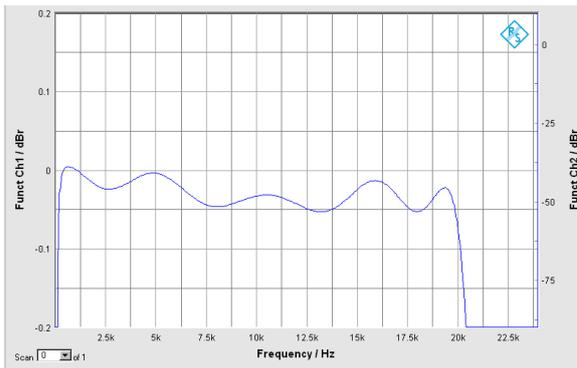


Abbildung 6.1.: Amplitudengang
DSK-Board

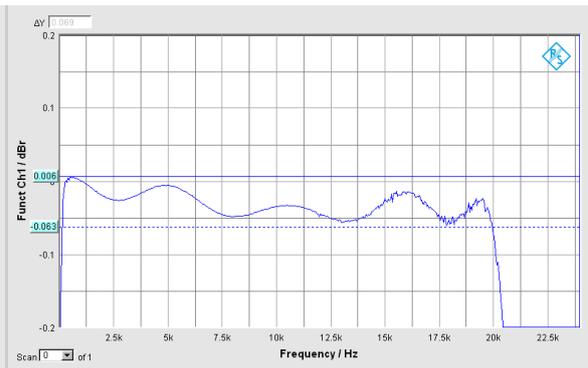


Abbildung 6.2.: Amplitudengang bei
 $v_B = 384$ kbps

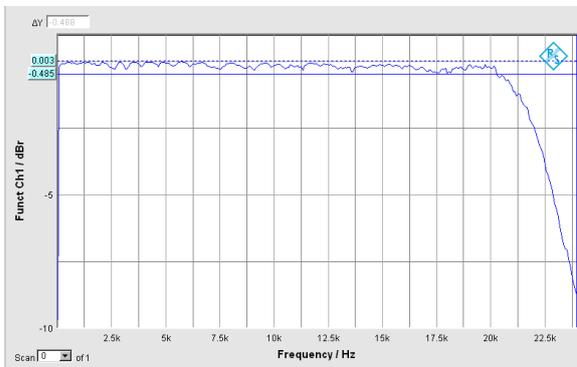


Abbildung 6.3.: Amplitudengang bei
 $v_B = 192$ kbps

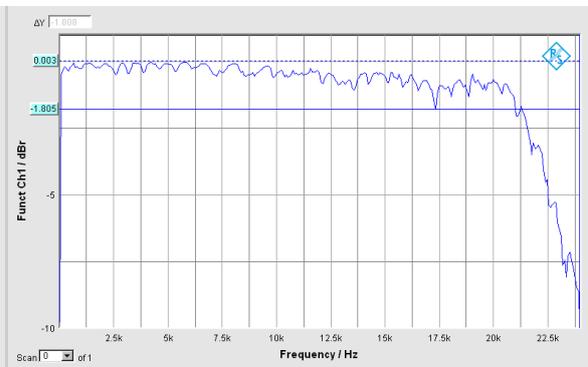


Abbildung 6.4.: Amplitudengang bei
 $v_B = 128$ kbps

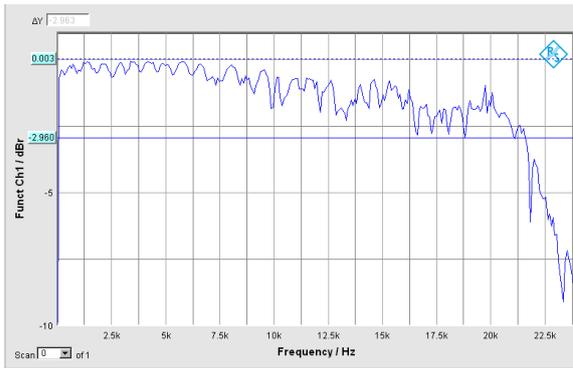


Abbildung 6.5.: Amplitudengang bei $v_B = 96$ kbps

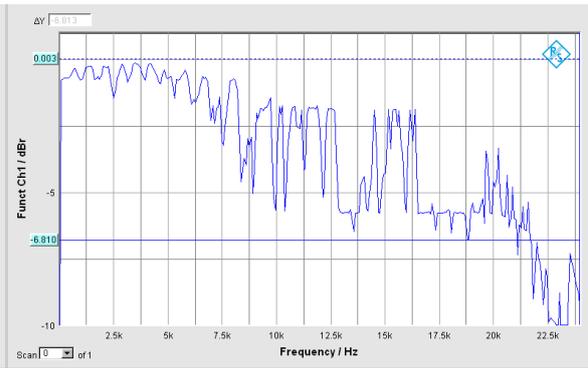


Abbildung 6.6.: Amplitudengang bei $v_B = 64$ kbps

6.3. Spektrogramme

Um die Spektrogramme aufnehmen und mit jenen aus der Simulation vergleichen zu können, wird die interne Soundkarte eines Desktop-PCs verwendet. Diese bietet mit 48 kHz Abtastfrequenz und 16 Bit Auflösung eine ausreichende Bitrate für die Signalanalyse. Der Messaufbau ist in Abbildung 6.7 dargestellt.

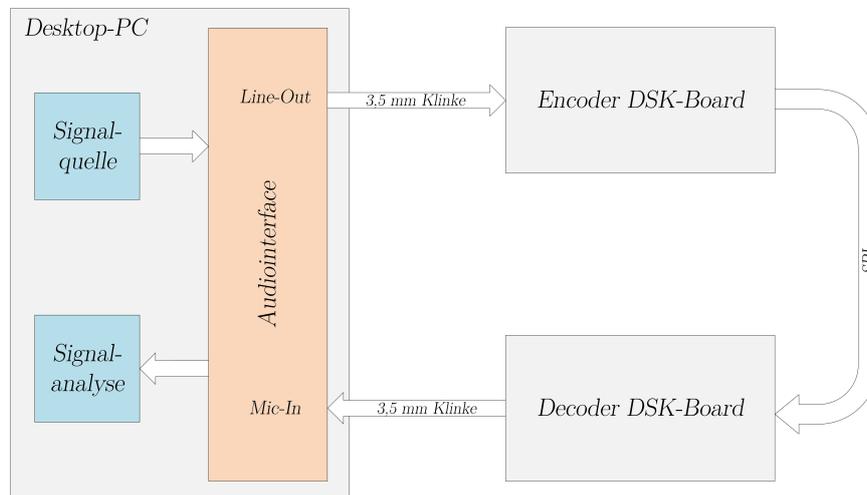


Abbildung 6.7.: Darstellung des Aufbaus zur Messung der Spektrogramme

Insgesamt zeigen die Spektrogramme kaum signifikante Unterschiede zur Simulation. Die Verluste der Komprimierung treten an gleichen Stellen auf. Dies fällt besonders bei $v_B=64$ kbps der Musikstücke \mathcal{S}^{Bra} und \mathcal{S}^{Joh} auf. Unstimmigkeiten sind lediglich in der Amplitude auszumachen, welche sich bei der Messung in Hardware stärker ausprägt. Die Unstimmigkeiten liegen mehr an den nicht-idealen Bauteilen der Messstrecke als an der abweichenden Umsetzung des Verfahrens zwischen Simulation und Hardware.

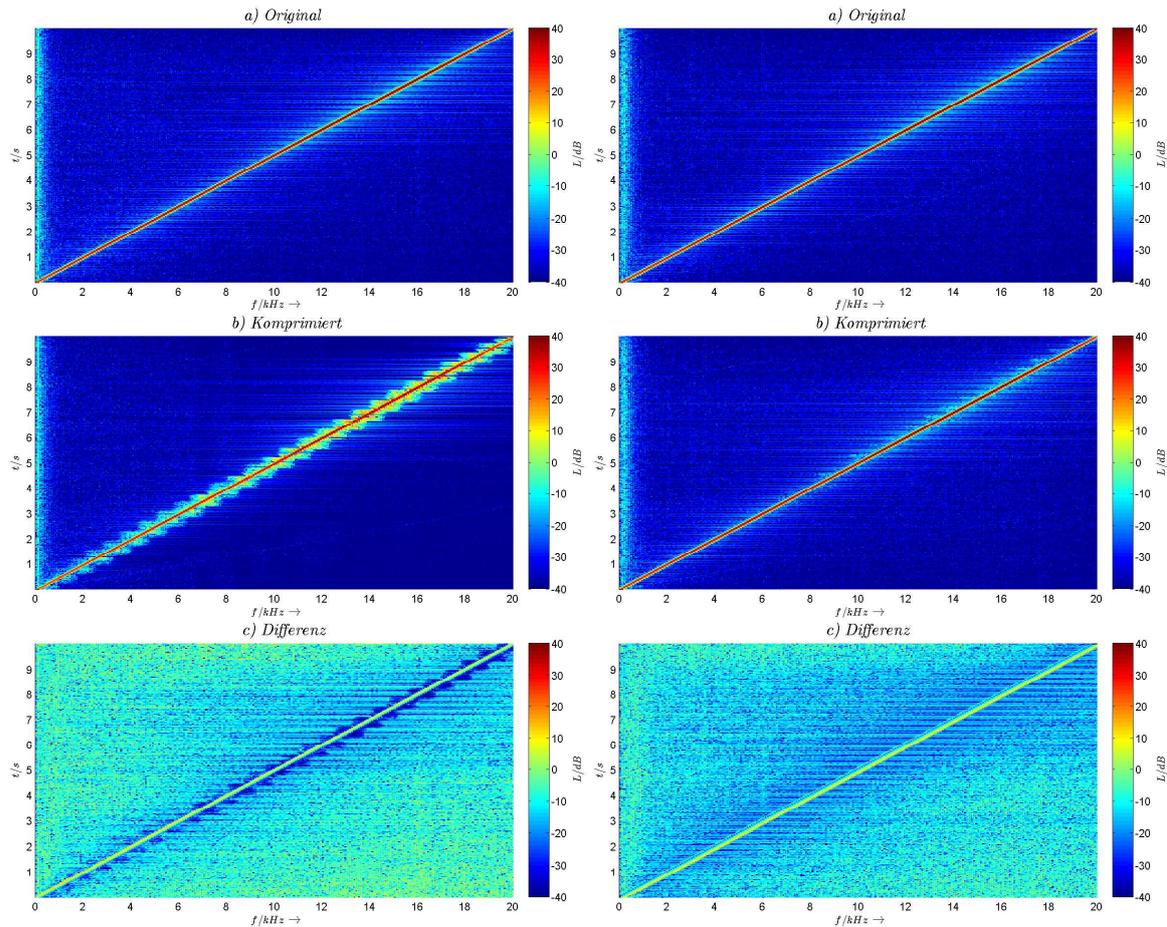


Abbildung 6.8.: Spektrogramm \mathcal{S}^{Swe} mit $v_B = 64$ kbps (Messung)

Abbildung 6.9.: Spektrogramm \mathcal{S}^{Swe} mit $v_B = 384$ kbps (Messung)

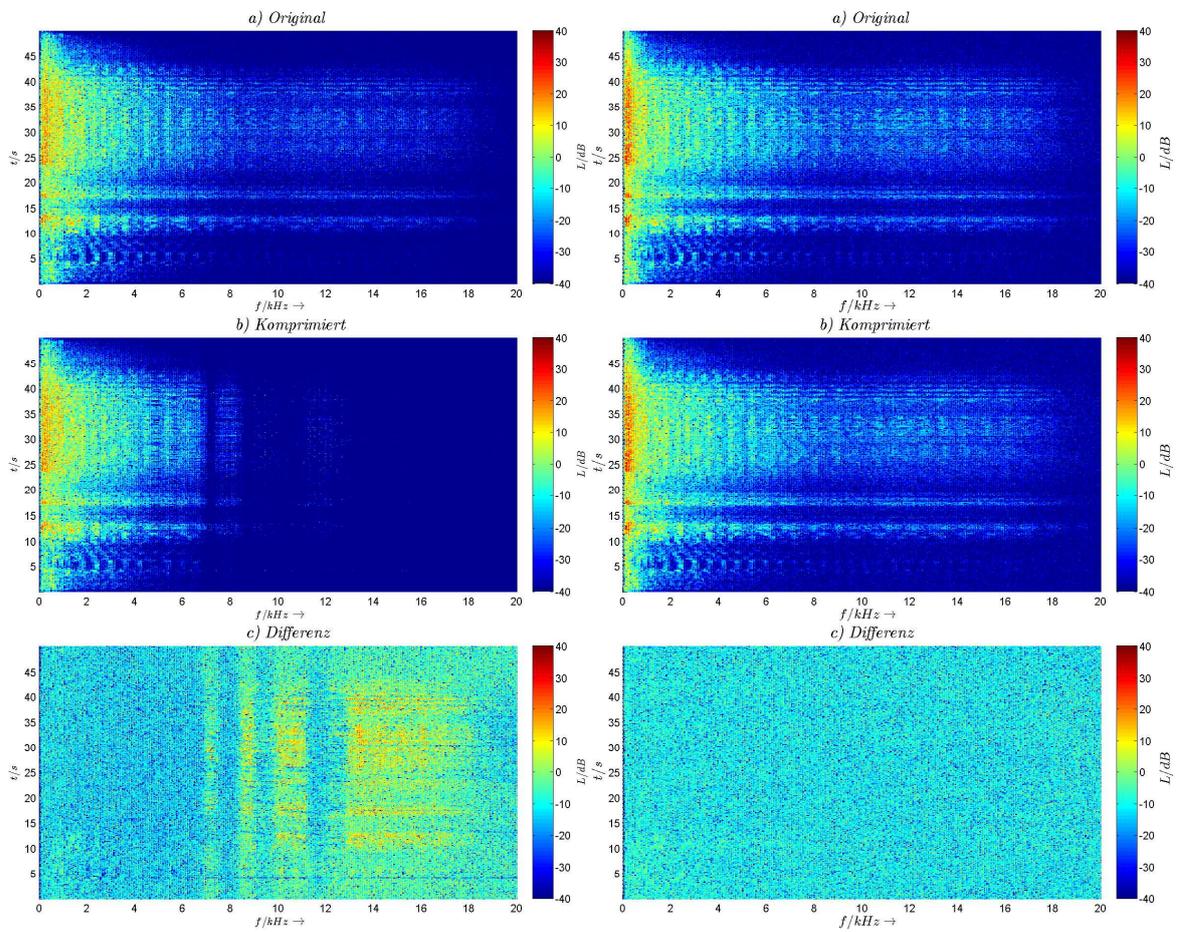


Abbildung 6.10.: Spektrogramm \mathcal{S}^{Bra}
mit $v_B = 64$ kbps
(Messung)

Abbildung 6.11.: Spektrogramm \mathcal{S}^{Bra}
mit $v_B = 384$ kbps
(Messung)

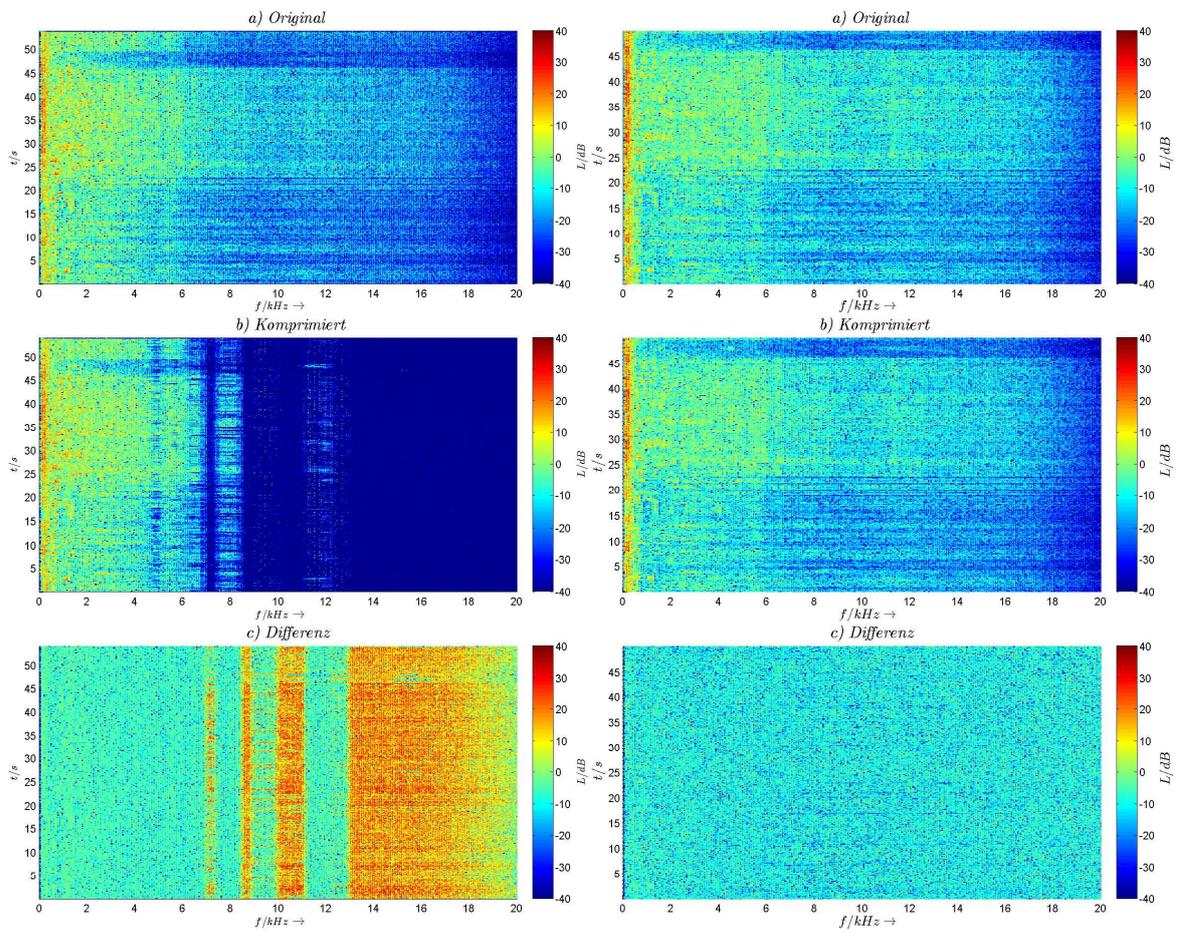


Abbildung 6.12.: Spektrogramm \mathcal{S}^{Joh}
mit $v_B = 64$ kbps
(Messung)

Abbildung 6.13.: Spektrogramm \mathcal{S}^{Joh}
mit $v_B = 384$ kbps
(Messung)

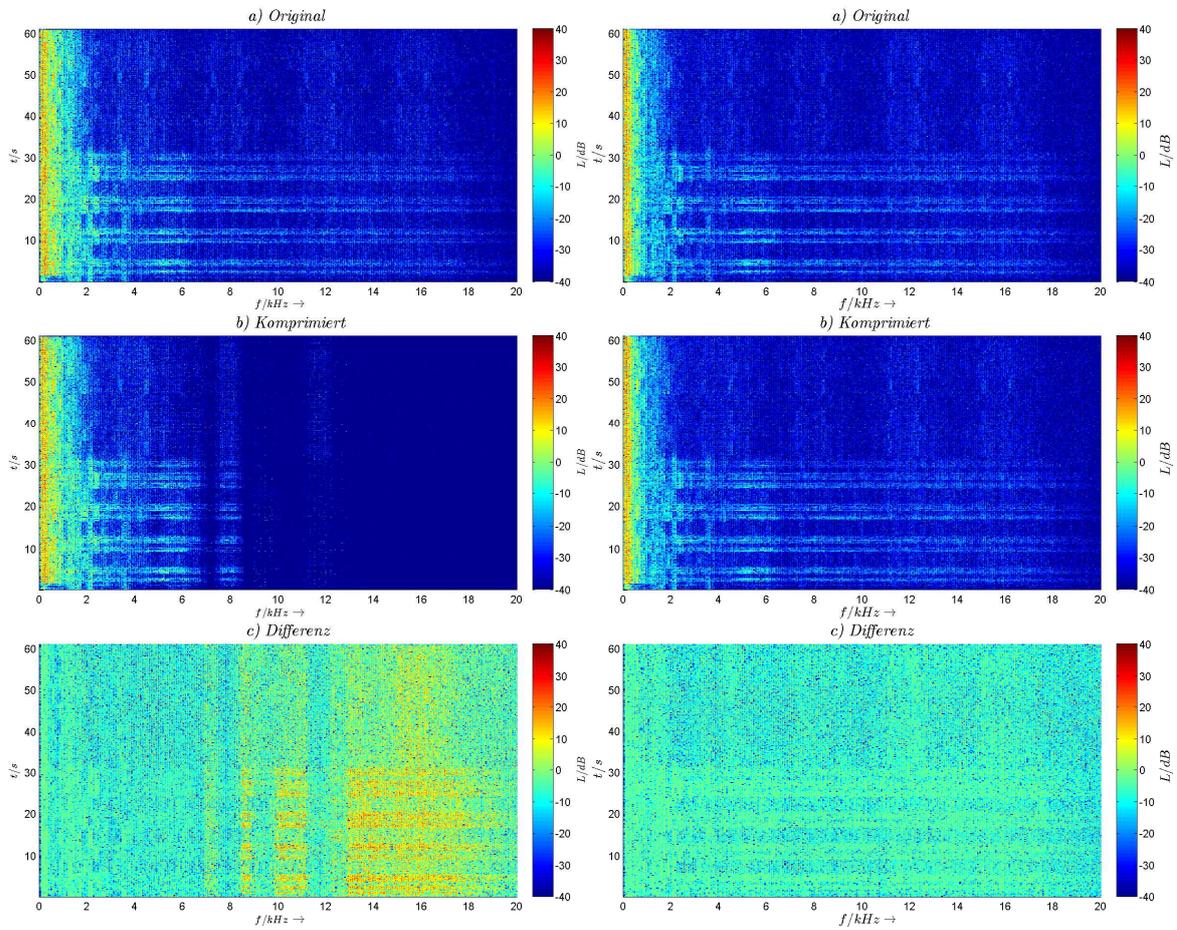


Abbildung 6.14.: Spektrogramm \mathcal{S}^{Kal}
mit $v_B = 64$ kbps
(Messung)

Abbildung 6.15.: Spektrogramm \mathcal{S}^{Kal}
mit $v_B = 384$ kbps
(Messung)

6.4. Hörtest

Einleitend ist zu erwähnen, dass es sich hier um keinen professionellen Hörtest handelt. Dies ist weder notwendig, noch aufgrund begrenzter Zeit, Anzahl an Testpersonen und verfügbarer Spezialgeräte möglich. Dennoch soll ein Hörtest mit einigen Testhörern die subjektiven Auswirkungen verschiedener Bitraten dem Leser dargestellt werden. Mit den vorliegenden Messungen und eigener Hörproben werden die drei Bitraten $v_B = 192$ kbps, $v_B = 128$ kbps und $v_B = 96$ kbps für den Test gewählt. Aufgrund der Spektrogramme und der Amplitudengängen ist zu erwarten, dass mit sinkender Bitrate zunehmend die Höhen gedämpft werden, was sich auf die Klangqualität auswirkt. Als Test-Methode wird jedem Hörer einmal der originale Ausschnitt des Musikstücks und einmal der Ausschnitt des komprimierten Musikstücks vorgespielt. Die Reihenfolge variiert dabei. Soweit es möglich war, sollte anschließend der Hörer entscheiden, welches der beiden Musikstücke bitreduziert und welches das originale war. Die Ergebnisse sind in Tabelle 6.2 aufgestellt. Die Entscheidungen bei $v_B = 192$ kbps zeigen, dass die Hörer überwiegend geraten haben müssen. Bei $v_B = 128$ kbps überwiegen die korrekt erkannten Aussagen nur knapp. Dies spricht immer noch für eine gute Klangqualität. Bei $v_B = 96$ kbps hingegen ist die Klangqualität so weit beeinträchtigt, dass fast jeder Hörer bei allen Musikstücken die bitreduzierte Variante erkannte. Die Spektrogramme zeigten, dass bei den Musikstücken \mathcal{S}^{Bra} und \mathcal{S}^{Joh} mit stärkerer Kompression deutlich mehr an Höhen verloren geht als es bei \mathcal{S}^{Kal} der Fall ist. Dies spiegelt sich auch etwas an den Hörtestergebnissen wieder. So wurden bei diesen beiden Musikstücken und $v_B = 128$ kbps mehr richtige Aussagen getroffen.

\mathcal{S}^{Bra}	$v_B = 192$ kbps	$v_B = 128$ kbps	$v_B = 96$ kbps
Testhörer 1	✓	✓	✓
Testhörer 2	<i>k.A.</i>	×	×
Testhörer 3	×	✓	✓
Testhörer 4	<i>k.A.</i>	×	✓
Testhörer 5	✓	✓	✓
Testhörer 6	✓	✓	✓
\mathcal{S}^{Joh}	$v_B = 192$ kbps	$v_B = 128$ kbps	$v_B = 96$ kbps
Testhörer 1	×	×	✓
Testhörer 2	×	✓	✓
Testhörer 3	✓	✓	✓
Testhörer 4	×	<i>k.A.</i>	✓
Testhörer 5	×	✓	✓
Testhörer 6	×	✓	✓
\mathcal{S}^{Kal}	$v_B = 192$ kbps	$v_B = 128$ kbps	$v_B = 96$ kbps
Testhörer 1	✓	✓	✓
Testhörer 2	×	×	✓
Testhörer 3	×	×	✓
Testhörer 4	<i>k.A.</i>	✓	✓
Testhörer 5	✓	×	✓
Testhörer 6	✓	✓	✓

Tabelle 6.2.: Ergebnisse des Hörtests

✓: Korrekt erkannt

×: Falsch erkannt

k.A.: Es konnte keine Aussage getroffen werden

Abschließend muss noch erwähnt werden, dass abweichend zum MPEG-1 Standard ein minimal gekürzter Rahmen verwendet wurde. So wurden die Zusatzinformationen des Headers (32 Bit) und der zyklischen Redundanzprüfung (16 Bit) nicht mit übertragen. Dadurch konnten mehr Nutzdaten bei der Bitzuweisung verwendet werden, was eventuell zu einer minimal besseren Audioqualität führen kann.

7. Zusammenfassung und Ausblick

In dieser Arbeit wurde ein Konzept zur Bitratenreduktion von Audiosignalen basierend auf dem MPEG-1 Verfahren entwickelt. Dies wurde anhand von Simulationen veranschaulicht und anschließend erfolgreich auf einer Hardwareplattform implementiert. Dabei konnte die Bearbeitung in Echtzeit realisiert werden.

Die Grundlagen dieses Verfahrens, welche auf psychoakustischen Effekten beruht, wurden analysiert und dargestellt. Ein wichtiger Bestandteil stellte die Cosinusmodulierte-Filterbank dar. Durch eine Optimierung konnten Verzerrungen und Aliasing deutlich verbessert werden. Obwohl keine perfekte Rekonstruktion erreicht wird, bietet diese Art der Filterbank für den Bereich Audio eine attraktive Lösung für die Zerlegung in Teilbänder.

Es wurde aufgezeigt, wie das psychoakustische Modell anhand von Mithörschwellen für das Gehör nicht wahrnehmbare Anteile kennzeichnet bzw. von hörbaren Anteilen separiert. Die Realisierung führte zunächst zu echtzeitkritischen Berechnungen. Durch geschickte Approximationen konnte jedoch deutlich Rechenleistung eingespart werden.

Die dynamische Bitzuweisung stellte das Bindeglied zwischen psychoakustischen Modell und Quantisierung dar. Es wurde aufgezeigt, wie sie bei vorgegebener Bitrate den Mithörschwellen-Rauschabstand optimiert.

Ein Vergleich der Spektrogramme zeigte, dass Differenzen zum MPEG-1 Standard quasi vernachlässigbar sind. Abschließende Hörtests zeigten auch noch bei einer Bitrate von 128 kbps gute Klangqualität.

Die gemessenen Performance-Zeiten zeigten, dass die separate Implementierung des Encoders und Decoders noch Raum für Berechnungen lässt. Sollten die Zeiten dennoch nicht ausreichen, kann auch eine andere Hardwareplattform mit höherer Taktfrequenz und FLOPS in Betracht gezogen werden. Aufbauend auf diese Arbeit sind folgende Themen denkbar:

- **Erweiterung auf MP3:**

Das MPEG-1 Layer III Verfahren erzielt auch noch bei Bitraten um 96 kbps eine gute Klangqualität. Für die Realisierung eines Encoders kann die vorliegende Filterbank übernommen werden. Bei dieser bedarf es eine weitere Zerlegung jedes Teilbands durch eine modifizierte diskrete Cosinus-Transformation. Für noch bessere Kompressionsergebnisse wird der Quantisierung eine Entropiecodierung in Form einer Huffman-Codierung nachgeschaltet. Des Weiteren kommt ein anderes psychoakustisches Modell zum Einsatz, welches die Daten einer 1024 Punkte FFT nutzt. Das Implementierungsverfahren ist in [Sta91] beschrieben.

- **Steganographie:**

In dieser Arbeit konnte gezeigt werden, dass bei einigen Bitraten die Verluste von unserem Gehör nicht oder nur schwach wahrgenommen werden. Die Datenreduzierung des Encoders kann genutzt werden, um geheime oder vertrauliche Informationen als getarnte Nutzinformation in die „Schein“-Daten (Trägerdaten) einzubetten. Dabei kann es sich beispielsweise um eine verborgene Sprachnachricht handeln. Ein normaler Decoder gibt nur die Trägerdaten wieder, während ein spezieller Decoder die versteckte Nachricht entschlüsselt.

A. Mathematische Zusammenhänge

A.1. Polyphasenrepräsentation

Allgemein ist

$$x[n] = \sum_{\lambda=0}^{M-1} x_{\lambda}^{(p)}[n] = \sum_{\lambda=0}^{M-1} x[n] \cdot w_M[n - \lambda] \quad (\text{A.1})$$

die Polyphasendarstellung im Zeitbereich [Fli93]. Dabei ist jedes Teilsignal $x_{\lambda}^{(p)}[n]$ eine von insgesamt M -Polyphasenkomponenten des Signals $x[n]$. Es gilt weiter für

$$w_M[n] = \frac{1}{M} \sum_{k=0}^{M-1} W_M^{kn} = \begin{cases} 1 & \text{für } n = mM, m \text{ ganzzahlig} \\ 0 & \text{sonst} \end{cases} \quad (\text{A.2})$$

wobei

$$W_M = e^{-j\frac{2\pi}{M}} \quad (\text{A.3})$$

den Modulationsoperator, oder auch Drehoperator (*engl. twiddle factor*) darstellt. Zur Veranschaulichung sei ein diskretes Signal $x[n]$ mit 9 Abtastwerten gegeben, vgl. Abbildung A.1 a). Für eine Polyphasenzerlegung in $M = 3$ Komponenten bzw. $\lambda = 0, 1, 2$ Polyphasen, ergibt sich nach Gleichung A.1

$$x[n] = x_0^{(p)}[n] + x_1^{(p)}[n] + x_2^{(p)}[n], \quad (\text{A.4})$$

vgl. Abbildung A.1 b), c) und d).

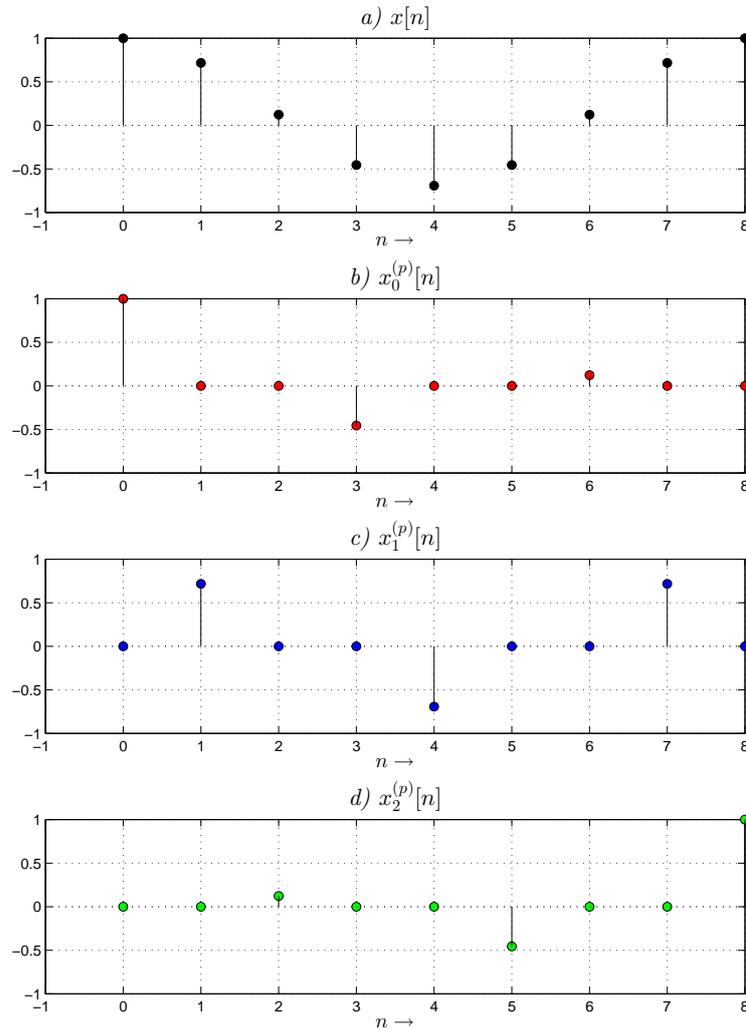


Abbildung A.1.: Darstellung der Polyphasen des Signals $x[n]$ für $M = 3$

Mittels z-Transformation folgt für die Polyphasendarstellung im Bildbereich

$$X(z) = \sum_{n=-\infty}^{\infty} x[n] \cdot z^{-n}. \quad (\text{A.5})$$

Das endliche Signal $x[n]$ kann mit Gleichung A.5 in der Form

$$\begin{aligned} X(z) = & x[0]z^{-0} + x[3]z^{-3} + x[6]z^{-6} \\ & + x[1]z^{-1} + x[4]z^{-4} + x[7]z^{-7} \\ & + x[2]z^{-2} + x[5]z^{-5} + x[8]z^{-8} \end{aligned} \quad (\text{A.6})$$

dargestellt werden. Nach Ausklammern der Potenzen $\lambda = 0 \dots 2$, für $M = 3$ ergibt sich

$$\begin{aligned} X(z) = & z^{-0} (x[0]z^{-0} + x[3]z^{-3} + x[6]z^{-6}) \\ & + z^{-1} (x[1]z^{-0} + x[4]z^{-3} + x[7]z^{-6}) \\ & + z^{-2} (x[2]z^{-0} + x[5]z^{-3} + x[8]z^{-6}). \end{aligned} \quad (\text{A.7})$$

In den runden Klammern resultieren ausschließlich Polynome, welche vielfache der Potenz $M = 3$ besitzen. Mit Gleichung A.5 und der Substitution $n = mM + \lambda$ erhält man den allgemeinen Ausdruck für die Polyphasendarstellung der z-Transformierten [Fli93]

$$X(z) = \sum_{\lambda=0}^{M-1} \sum_{m=-\infty}^{\infty} x[mM + \lambda] \cdot z^{-mM+\lambda} = \sum_{\lambda=0}^{M-1} z^{-\lambda} X_{\lambda}^{(p)}(z^M) \quad (\text{A.8})$$

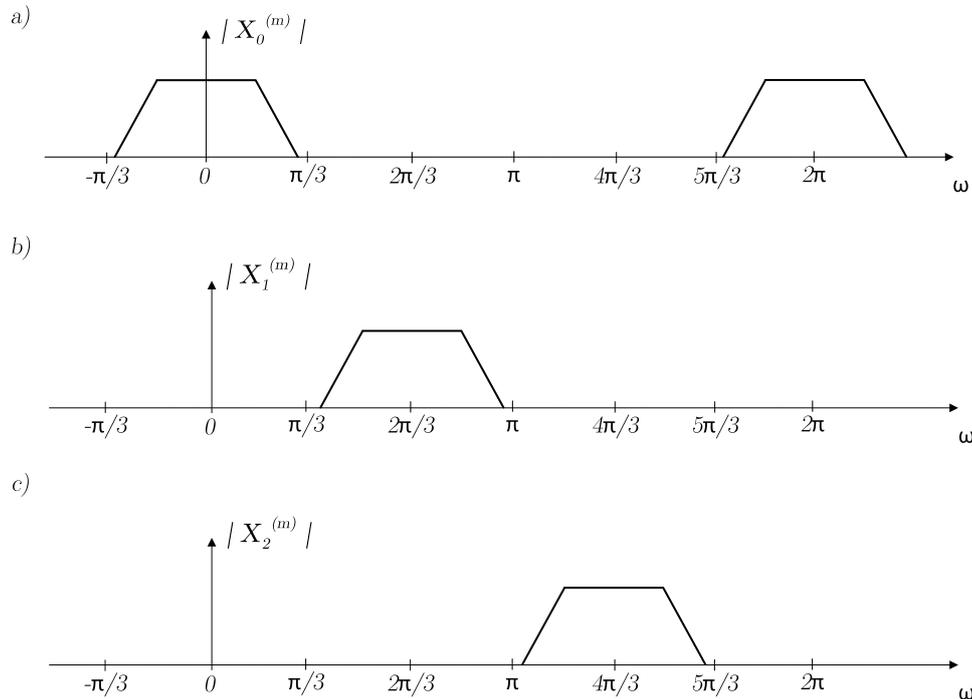
mit

$$X_{\lambda}^{(p)}(z^M) = \sum_{m=-\infty}^{\infty} x[mM + \lambda] \cdot z^{-mM}. \quad (\text{A.9})$$

Gleichung A.9 zeigt die allgemeine z-Transformierte der Polyphasenkomponenten, des Polynoms in eckigen Klammern aus Gleichung A.7. Mit dem Übergang zur zeitdiskreten Fouriertransformation $z \rightarrow e^{j\omega}$, wird deutlich, dass das Ausgangssignalspektrum $X(z) \rightarrow X(e^{j\omega})$ moduliert wird.

$$X^{(m)}(e^{j\omega}) = \sum_{k=0}^{M-1} X_k^{(m)}(e^{j\omega} \cdot e^{-j2\pi k/M}) = \sum_{k=0}^{M-1} X_k^{(m)}(e^{j\omega} \cdot W_M^k) \quad (\text{A.10})$$

Man spricht daher auch von der sogenannten Modulationsdarstellung [Fli93]. Abbildung A.2 verdeutlicht dies.

Abbildung A.2.: Modulationsdarstellung für $M=3$

A.2. Dezimation und Interpolation

Unter einer Dezimation versteht man die Herabsetzung der Abtastrate (Abwärtstastung) eines diskreten Signals mit vorgeschaltetem Anti-Aliasing Filter. Für die Abwärtstastung um den Faktor M gilt im Zeitbereich allgemein

$$y[m] = x[m \cdot M]. \quad (\text{A.11})$$

Das Ausgangssignal $y[m]$ besitzt nach der Abwärtstastung um den Faktor M lediglich noch jeden M -ten Abtastwert des ursprünglichen Signals $x[m]$. Das Spektrum von Gleichung A.11 lässt sich mit Hilfe von Modulationskomponenten ausdrücken [Fli93]. Mit der z -Transformation erhält man den Ausdruck

$$Y(z^M) = \frac{1}{M} \sum_{k=0}^{M-1} X(zW_M^k), \quad (\text{A.12})$$

mit

$$W_M = e^{-j\frac{2\pi}{M}}. \quad (\text{A.13})$$

Im Spektrum tauchen die mit W_M^k modulierten Spektren auf. Abbildung A.3 verdeutlicht dieses für $M = 3$. Zur Vermeidung von Aliasing ist es ersichtlich, dass die Bandbreite des zu dezimierenden Signals auf π/M bzw. $F_s/2M$ begrenzt werden muss. Die Kaskadierung von Anti-Aliasing Filter und Abwärtstaster wird auch als Dezimator bezeichnet [Sau07].

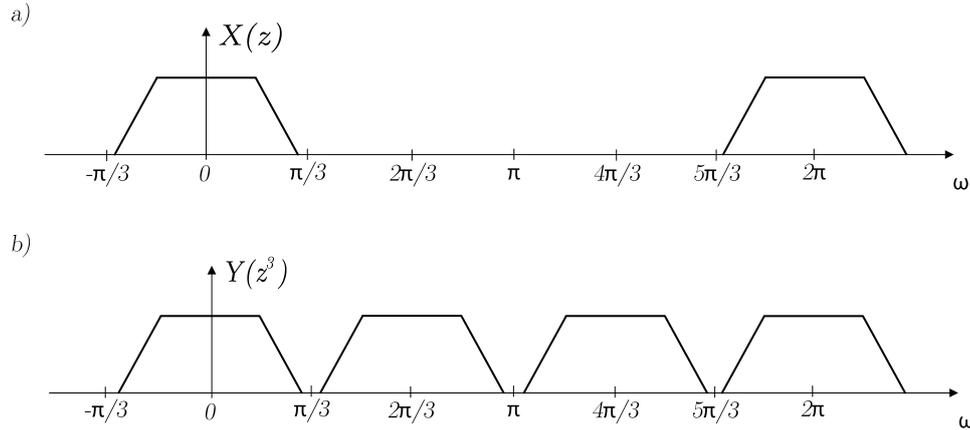


Abbildung A.3.: Spektren bei Dezimation für $M=3$

Das Gegenstück zur Dezimation ist die Interpolation, welche aus einer Erhöhung der Abtastrate (Aufwärtstastung) und nachgeschaltetem Anti-Imaging Filter besteht. Die Aufwärtstastung um den Faktor L lautet für den Zeitbereich allgemein

$$y[n] = \begin{cases} x[n/L] & \text{für } n = mL, m \text{ ganzzahlig} \\ 0 & \text{sonst} \end{cases} \quad (\text{A.14})$$

Die z-Transformation von Gleichung A.14 liefert den Ausdruck

$$Y(z) = X(z^L). \quad (\text{A.15})$$

Die Spektren bei vielfachen der niedrigeren Abtastfrequenz tauchen als Spiegelfrequenzen der neuen-, höheren-Abtastfrequenz auf. Diese müssen durch ein Filter, dem Anti-Image-Filter, hinreichend gedämpft werden. Abbildung A.4 veranschaulicht dieses für $L = 3$.

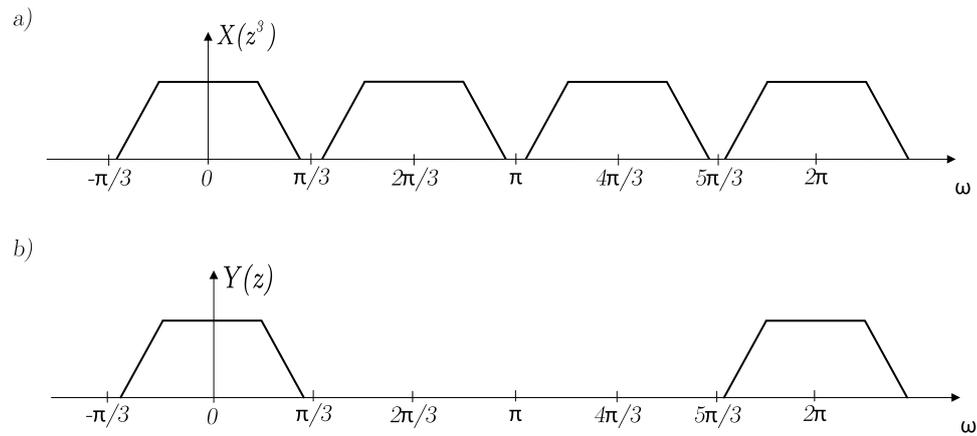


Abbildung A.4.: Spektren bei Interpolation für $M=3$

B. Spektrogramme

B.1. Simulation

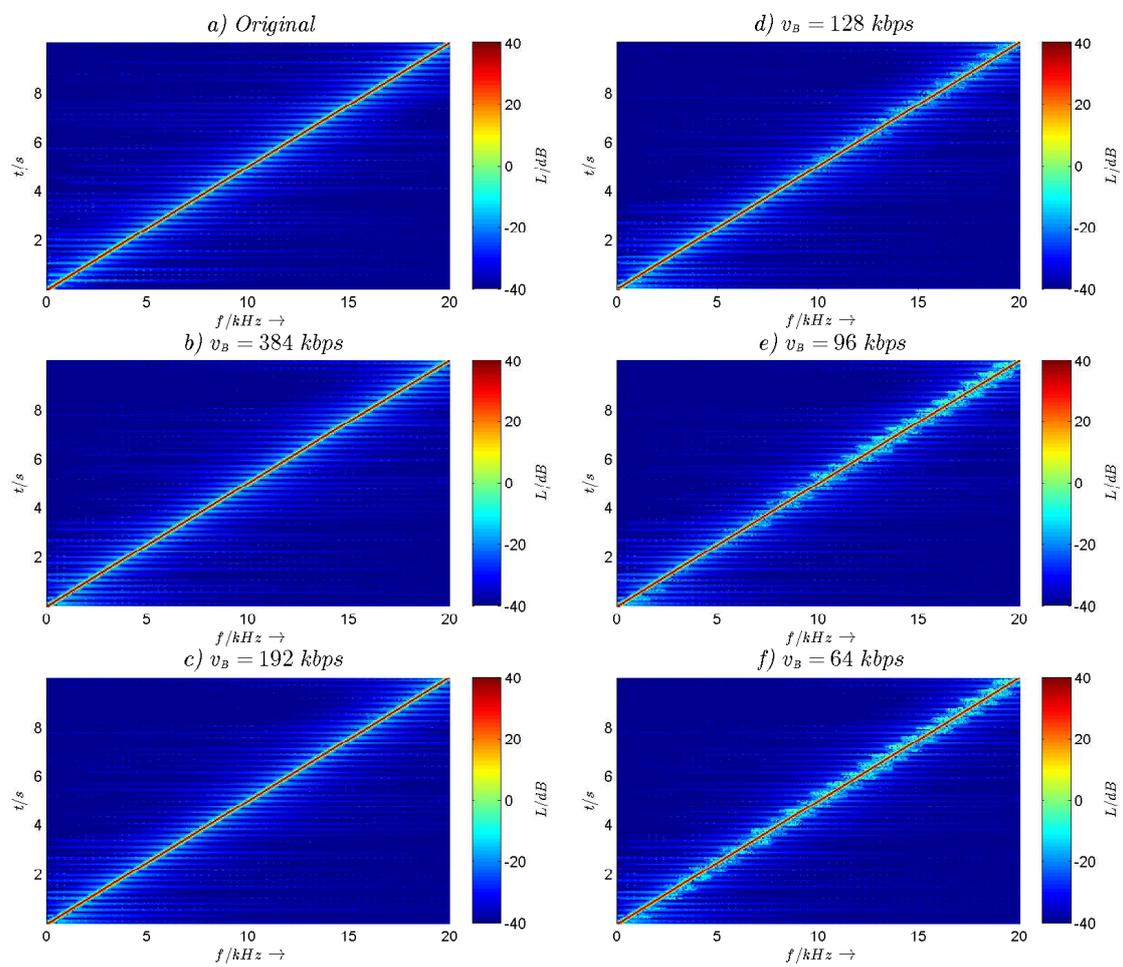


Abbildung B.1.: Spektrogramme \mathcal{S}^{Swe} bei allen Bitraten (Simulation)

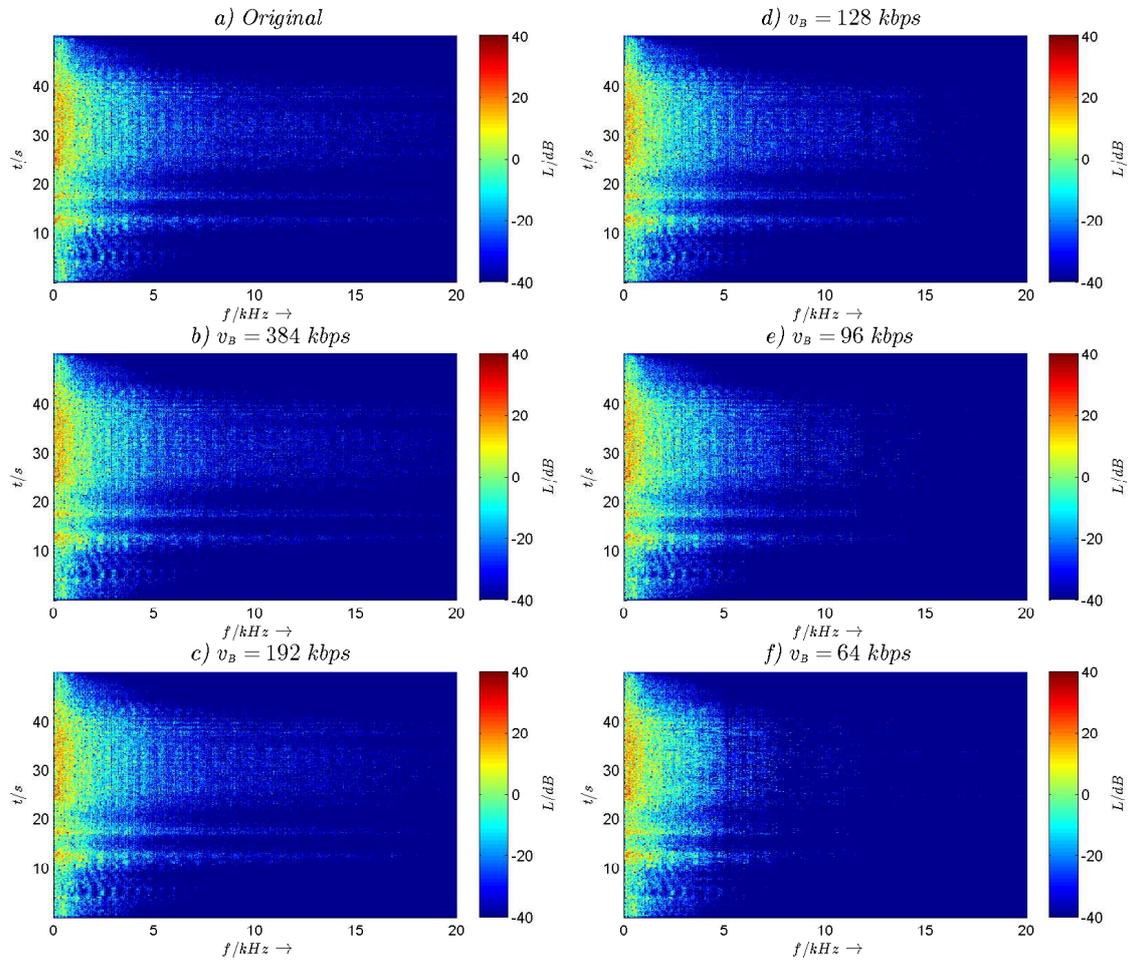


Abbildung B.2.: Spektrogramme \mathcal{S}^{Bra} bei allen Bitraten (Simulation)

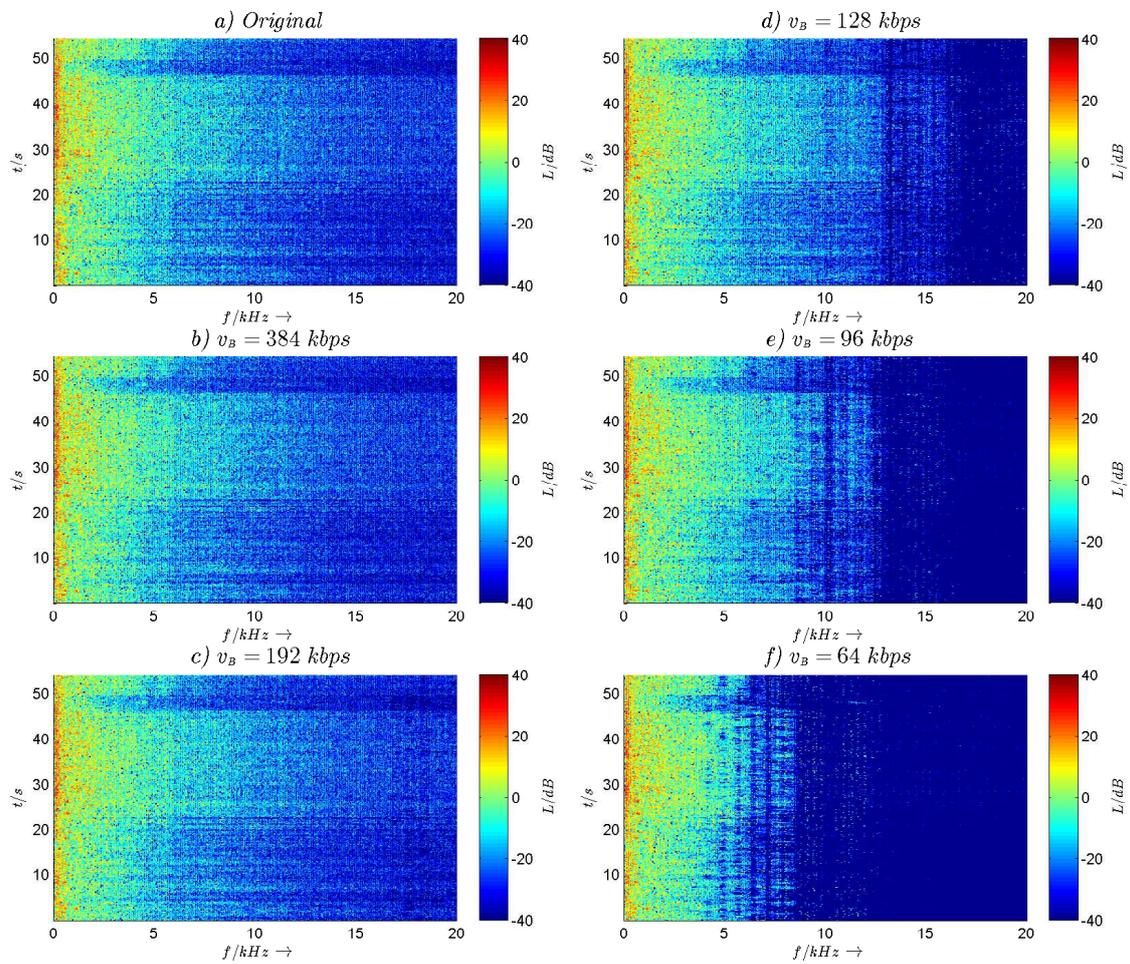


Abbildung B.3.: Spektrogramme \mathcal{S}^{Joh} bei allen Bitraten (Simulation)

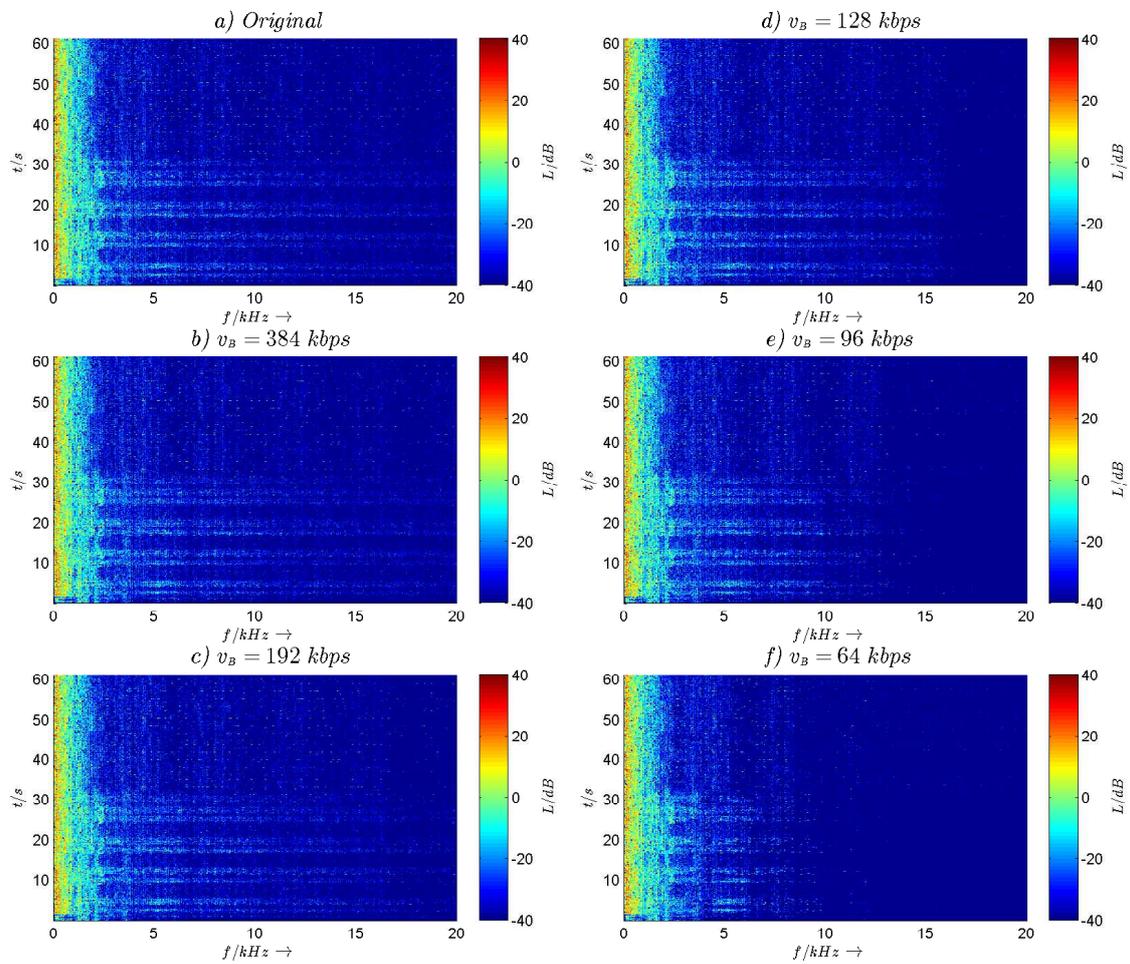
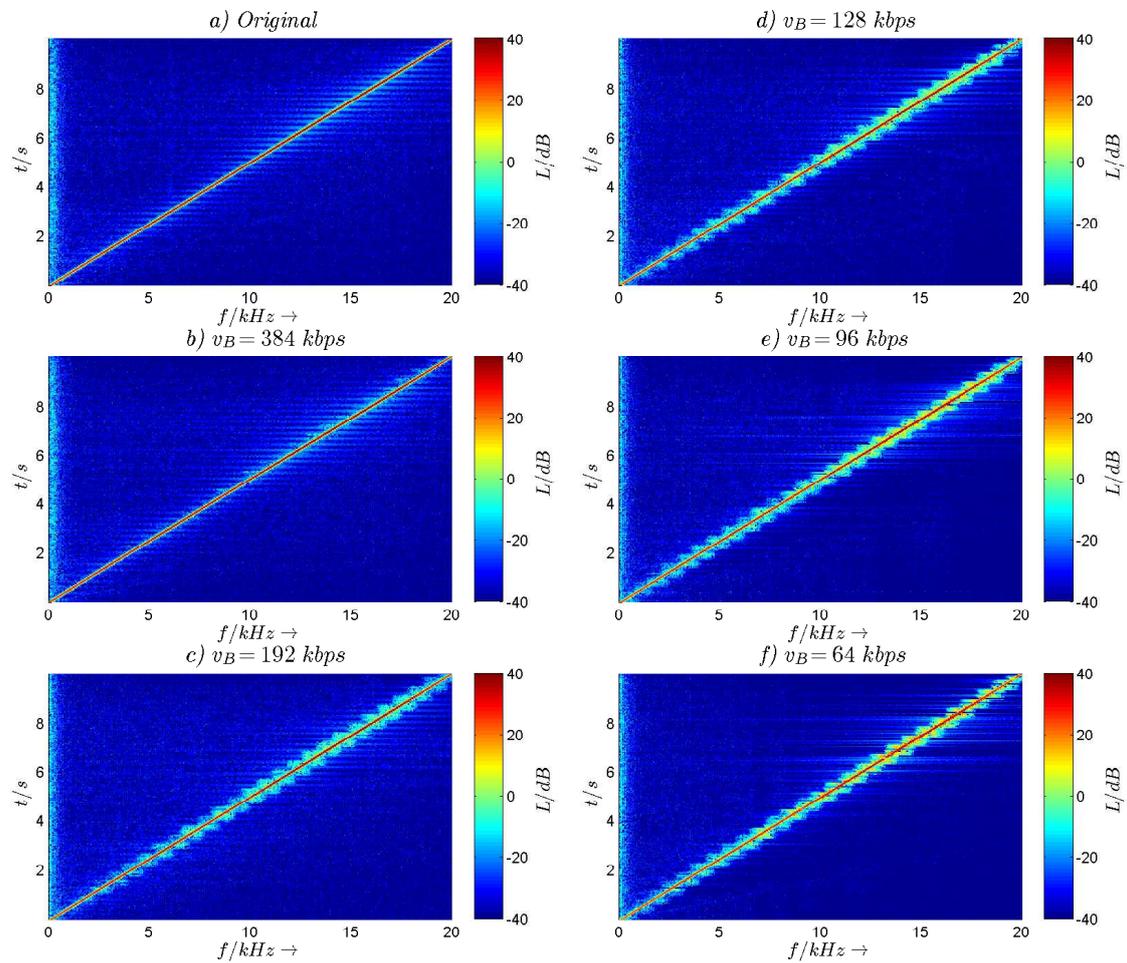


Abbildung B.4.: Spektrogramme \mathcal{S}^{Kal} bei allen Bitraten (Simulation)

B.2. DSP Messungen

Abbildung B.5.: Spektrogramme $\mathcal{S}^{S^{we}}$ bei allen Bitraten (Messung)

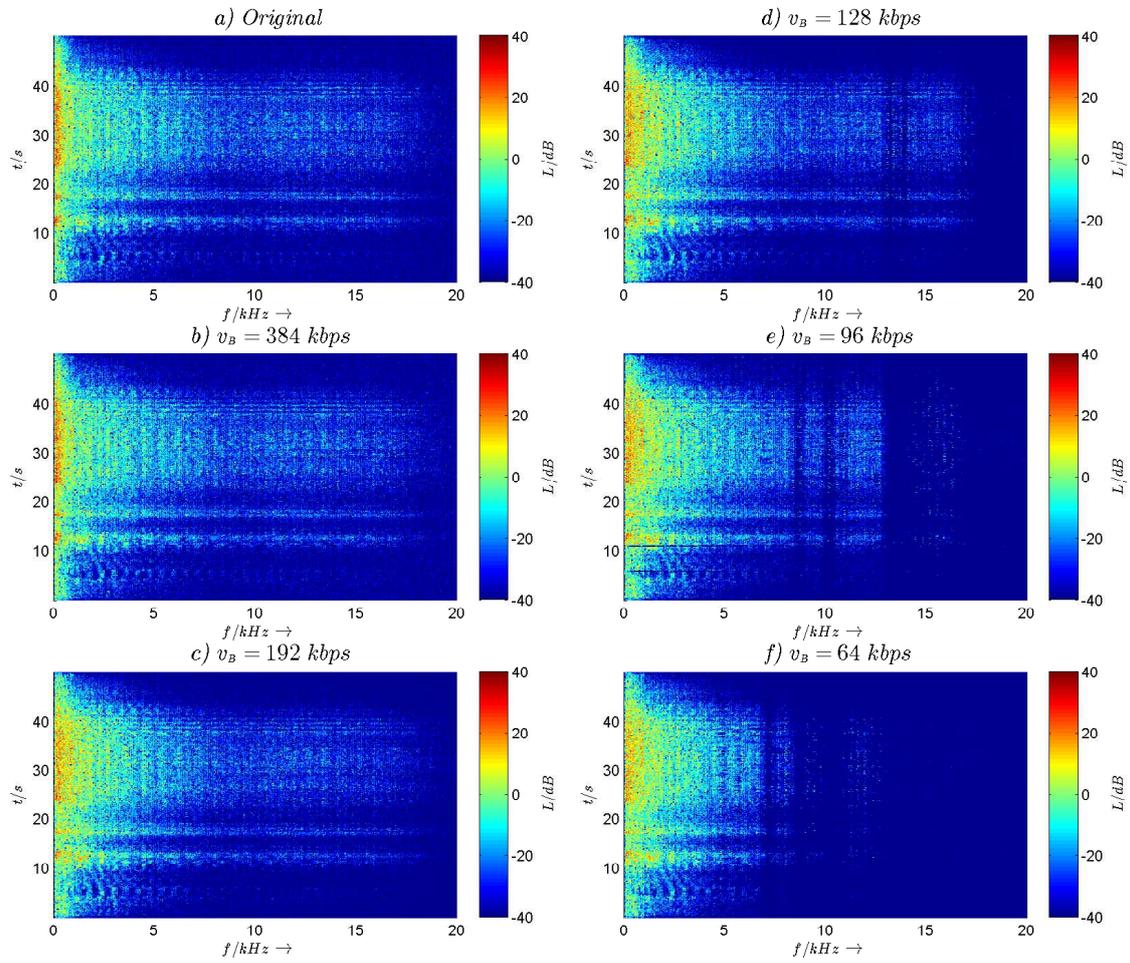


Abbildung B.6.: Spektrogramme \mathcal{S}^{Bra} bei allen Bitraten (Messung)

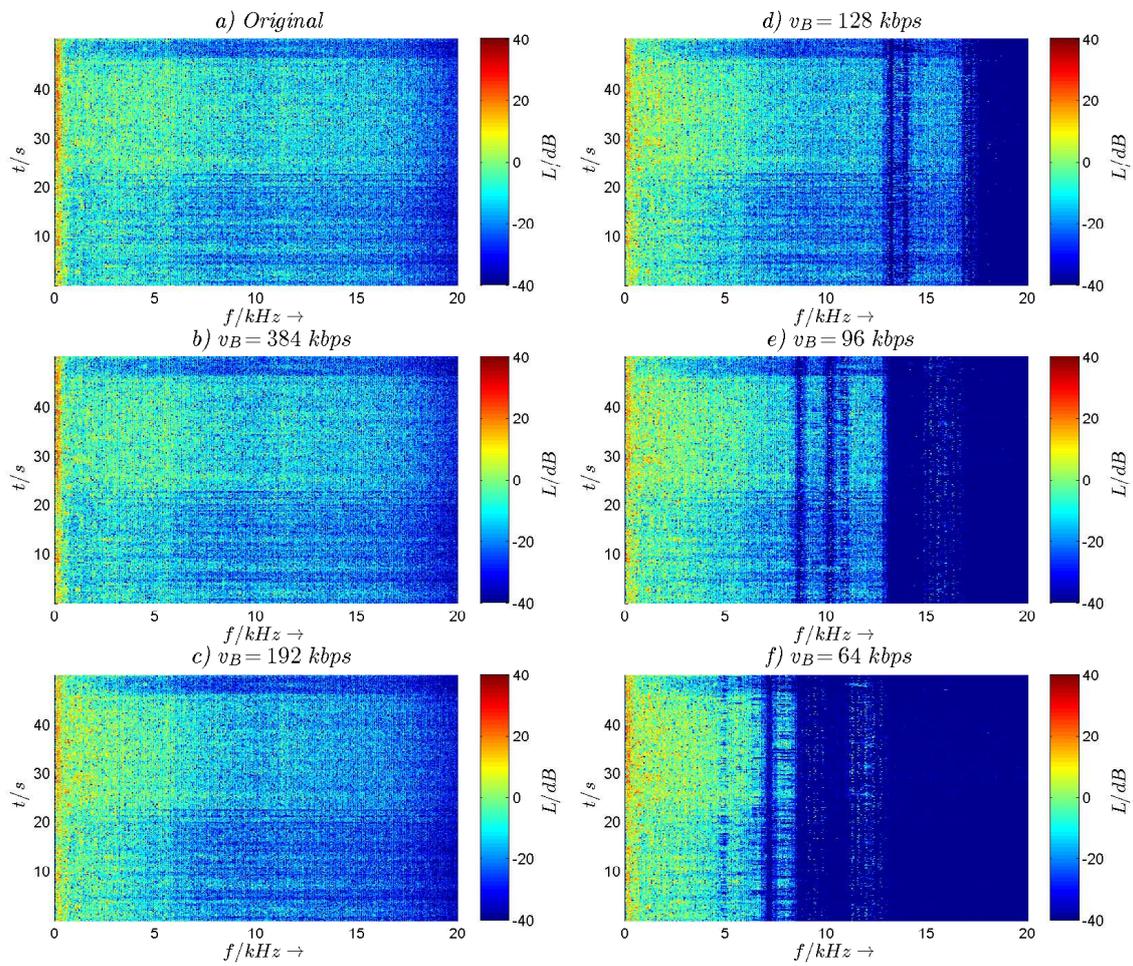


Abbildung B.7.: Spektrogramme \mathcal{S}^{Joh} bei allen Bitraten (Messung)

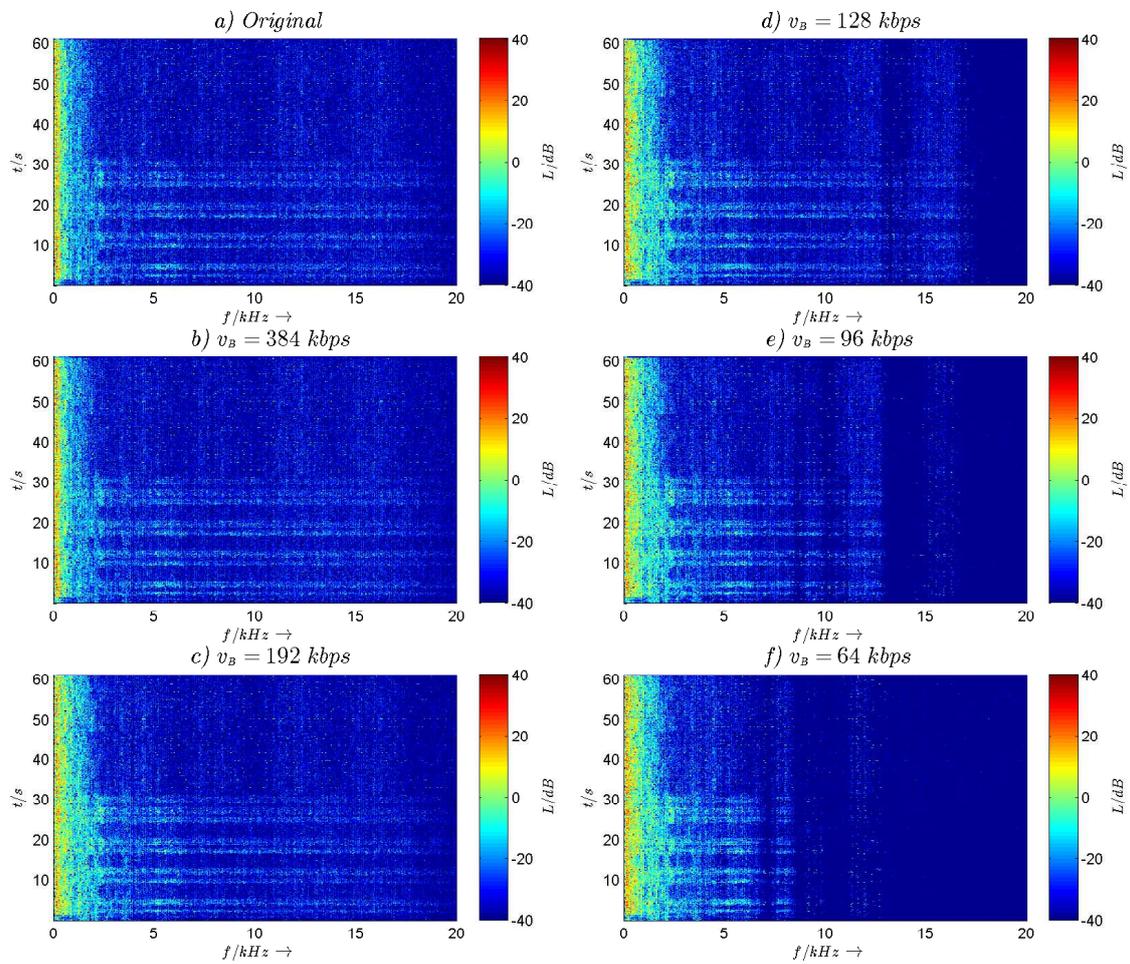


Abbildung B.8.: Spektrogramme S^{Kal} bei allen Bitraten (Messung)

C. Inhalt des Datenträgers

Der beigefügte Datenträger enthält die folgenden drei Ordner mit dem Inhalt:

1. **/01_Masterarbeit_Kim_Radmacher**
Dieser Ordner enthält die vorliegende Masterarbeit als pdf-Datei
2. **/02_MATLAB**
Dieser Ordner enthält die in dieser Arbeit entwickelten *MATLAB*-Skripte
 - a. **/Filterbank:** Simulationen der beschriebenen Filterbänke
 - b. **/show_mpeg_method:** Darstellung des MPEG-1 Verfahrens bzw. alle abgebildeten MPEG-1 Simulationen
 - c. **/compress_audio:** MATLAB-basierte Komprimierung von Audio-Dateien im .wav-Format
3. **/03_CCS_Projekte**
Dieser Ordner enthält die implementierten CCS-Projekte
 - a. **/I2S_32Bit_Messung:** Amplitudengang-Messung der Analyse-Filterbank
 - b. **/Filterbank:** Messung der Filterbank-Verzerrungsfunktion
 - c. **/EncodeDecode:** MPEG-1 Encoder und Decoder realisiert auf einem DSK-Board
 - d. **/MPEG_Encoder:** MPEG-1 Encoder für Realisierung auf zwei DSK-Boards
 - e. **/MPEG_Decoder:** MPEG-1 Decoder für Realisierung auf zwei DSK-Boards

Literaturverzeichnis

- [CR10] CHASSAING, Rulph ; REAY, Donald: *Digital Signal Processing and Applications with the TMS320C6713 and TMS320C6416 DSK*. 2. Auflage. Wiley-IEEE Press, 2010. – ISBN 9780470138663
- [Fli93] FLIEGE, Norbert: *Multiraten Signalverarbeitung Theorie und Anwendungen*. Teubner Verlag, 1993. – ISBN 9783519061557
- [Gro14] GROUP, Moving Picture E.: *MPEG Homepage*. <http://mpeg.chiariglione.org>, 2014. – [Abruf 05.05.2015]
- [Int15] INTERNATIONAL, Iowegian: *Digital Signal Processing Tricks*. <http://dspguru.com/dsp/tricks/>, 2015. – [Abruf 05.06.2015]
- [Kam11] KAMMEYER, Karl-Dirk: *Nachrichtenübertragung*. 5. Auflage. Vieweg+Teubner Verlag, 2011. – ISBN 9783834808967
- [MA97] MURPHY, C.D. ; ANANDAKUMAR, K.: *Real-time MPEG-1 audio coding and decoding on a DSP chip*. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=580383&queryText=Real-time+MPEG-1+audio+coding+and+decoding+on+a+DSP+chip&newsearch=true&searchField=Search_All, 1997. – [Abruf 16.06.2015]
- [Mit02] MITRA, Sanjit K.: *Digital Signal Processing: A Computer-Based Approach*. 2. Auflage. McGraw-Hill Inc.,US, 2002. – ISBN 9780071226073
- [MV13] MAES, Jan ; VERCAMMEN, Marc: *Digital Audio Technology*. 4. Auflage. Focal Press, 2013. – ISBN 9780240516547
- [Sal06] SALOMON, David: *Data Compression: The Complete Reference*. 4. Auflage. Springer, 2006. – ISBN 9781846286025
- [Sau04] SAUVAGERD, Ulrich: *Vorlesungsskript Digitale Audiosignalverarbeitung*. 2004
- [Sau07] SAUVAGERD, Ulrich: *Vorlesungsskript Digitale Signalverarbeitung auf Signalprozessoren*. 2007
- [SPA07] SPANIAS, Andreas ; PAINTER, Ted ; ATTI, Venkatraman: *Audio Signal Processing and Coding*. Wiley-Interscience, 2007. – ISBN 9780471791478

- [Sta91] STANDARDIZATION, ISO International O.: *CD 11172-3 Coding of Moving pictures and associated audio for digital storage media at up to about 1.5 mbit/s part 3 audio*. <http://www.eece.colorado.edu/~fmeyer/class/ecen4532/mpeg1-layerIII.pdf>, 1991. – [Abruf 16.04.2015]
- [Str09] STRUTZ, Tilo: *Bilddatenkompression: Grundlagen, Codierung, Wavelets, JPEG, MPEG, H.264*. 4. Auflage. Vieweg+Teubner Verlag, 2009. – ISBN 9783834804723
- [Tex04] TEXAS INSTRUMENTS TI: *TLV320AIC23B Stereo Audio Codec*. <http://www.ti.com/lit/ds/symlink/tlv320aic23b.pdf>, 2004. – [Abruf 03.06.2015]
- [Tex06a] TEXAS INSTRUMENTS TI: *MS320C6000 DSP Enhanced Direct Memory Access (EDMA) Controller Reference Guide*. <http://www.ti.com/lit/ug/spru234c/spru234c.pdf>, 2006. – [Abruf 16.04.2015]
- [Tex06b] TEXAS INSTRUMENTS TI: *TMS320C6000 DSP Multichannel Buffered Serial Port (McBSP) Reference Guide*. <http://www.ti.com/lit/ug/spru580g/spru580g.pdf>, 2006. – [Abruf 16.04.2015]
- [Tex06c] TEXAS INSTRUMENTS TI: *TMS320C6713B Floating-Point Digital Signal Processor Device-Specific Documentation*. <http://www.ti.com/lit/ds/symlink/tms320c6713b.pdf>, 2006. – [Abruf 16.04.2015]
- [Tex06d] TEXAS INSTRUMENTS TI: *TMS320C67x/C67x+ DSP CPU and Instruction Set Reference Guide*. <http://www.ti.com/lit/ug/spru733a/spru733a.pdf>, 2006. – [Abruf 05.06.2015]
- [Tex14] TEXAS INSTRUMENTS TI: *TMS320C6000 McBSP: I2S Interface*. <http://www.ti.com/lit/an/spra595/spra595.pdf>, 2014. – [Abruf 05.06.2015]
- [UKW] UKW/TV-ARBEITSKREIS E.V.: *DAB Sender-Tabelle Hamburg*. <http://www.ukwtv.de/cms/deutschland-dab/hamburg-dab.html>, . – [Abruf 05.05.2015]
- [Vai92] VAIDYANATHAN, P. P.: *Multirate Systems and Filter Banks*. Prentice Hall, 1992. – ISBN 9780136057185
- [WS04] WALKER, William ; STRANNEBY, Dag: *Digital Signal Processing and Applications*. 2. Auflage. Newnes, 2004. – ISBN 0750663448
- [Zö8] ZÖLZER, Udo: *Digital Audio Signal Processing*. 2. Auflage. Wiley, 2008. – ISBN 9780470997857
- [ZF99] ZWICKER, Eberhard ; FASTL, Hugo: *Psychoacoustics: Facts and Models*. 2. Auflage. Springer, 1999. – ISBN 9783540650638

Versicherung über die Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen habe ich unter Angabe der Quellen kenntlich gemacht.

Hamburg, 29.06.2015

Ort, Datum

Unterschrift