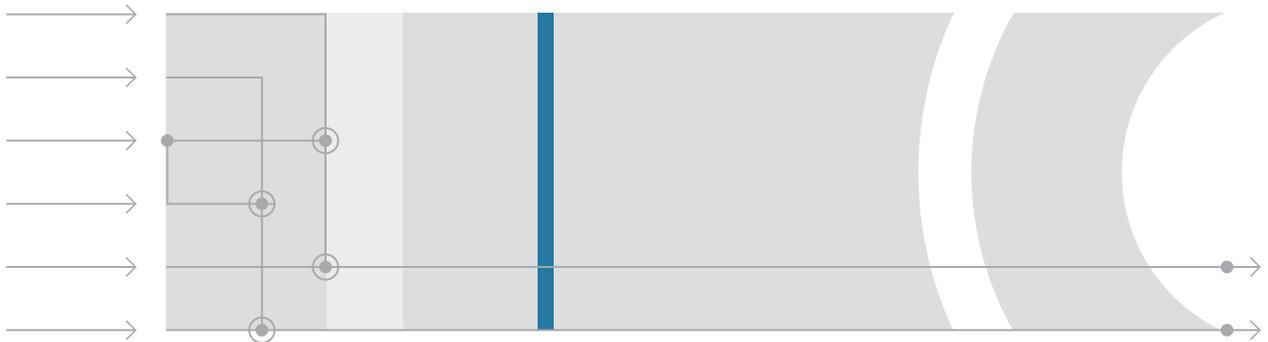


Institut für Rundfunktechnik  
Audiosystemtechnik

Hochschule der Medien Stuttgart  
Fakultät Electronic Media  
Studiengang Audiovisuelle Medien

## Automatischer Stereo-Downmix von 5.1-Mehrkanalproduktionen

Optimierung des IRT-Downmix hinsichtlich  
subjektiver und objektiver Qualitätskriterien



Diplomarbeit vorgelegt von  
Bernfried Runow

Erstprüfer Prof. Oliver Curdt, HdM  
Zweitprüfer Dipl. Ing. Gerhard Stoll, IRT

Stuttgart, den 6. Juli 2008

## Vorwort

Die Audioproduktion stellt in der heutigen Zeit sowohl Tonmeister als auch Produzenten vor immer neue Herausforderungen: Eine effektive und künstlerische Arbeitsweise bei gleichzeitiger kostenreduzierter Bedienung mehrerer Endformate. Wurden Produktionen des vergangenen Jahrhunderts speziell auf ein Medium zugeschnitten, ergeben sich heute aus der Vielzahl der Wiedergabeformate und Einsatzmöglichkeiten zunehmend Schwierigkeiten in Bezug auf die Kompatibilität.

Automatisierte genreangepasste Algorithmen, die im Hintergrund verlässlich phasen- und amplitudenstabile Mischvorgänge übernehmen, sind heute unverzichtbare Tools für Rundfunk- und Fernsehtonübertragung sowie die Verbreitung über das Internet.

Der in dieser Arbeit vorgestellte Downmixalgorithmus stellt aktuellen Produktionen ein zuverlässiges Werkzeug zur Verfügung.

Silvio Naumann, Tonmeister für Audiovisuelle Medien





## Erklärung

Hiermit erkläre ich, meine Diplomarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel herangezogen zu haben.

Ich versichere, dass ich bisher keine Prüfungsarbeit mit gleichem oder ähnlichem Thema bei einer Prüfungsbehörde oder einer anderen Hochschule vorgelegt habe.

Ort, Datum

Bernfried Runow

# Inhalt

1	Einleitung.....	1
2	Mehrkanalton.....	3
2.1	Tonformate .....	3
2.1.1	Monofonie .....	4
2.1.2	Zweikanal-Stereofonie .....	4
2.1.3	5.1-Mehrkanalton.....	5
2.2	Mehrkanal beim Rundfunk.....	7
2.2.1	Produktion von Mehrkanalton.....	7
2.2.2	Tonträger .....	7
2.2.3	Möglichkeiten der Übertragung.....	8
2.3	Notwendigkeit des Downmixes .....	9
3	Downmix .....	11
3.1	Handmix.....	11
3.2	Automatische passive Downmixverfahren .....	11
3.2.1	ITU Downmix.....	11
3.2.2	Dolby Pro Logic II.....	12
3.3	Automatische aktive Downmixverfahren.....	13
3.3.1	Lexicon Logic7 .....	13
3.3.2	Coding Technologies .....	14
3.4	Ergebnisse vorangegangener Untersuchungen am IRT.....	15
3.4.1	Erster Hörversuch zum Thema Downmix am IRT.....	15
	Ergebnis des Hörversuchs.....	15
3.4.2	Zweiter Hörversuch zum Thema Downmix am IRT.....	16
4	Technische Grundlagen und Begriffe.....	17
4.1	Korrelation .....	17
4.2	Kohärenz.....	17
4.3	Addition von Amplituden und Pegeln .....	18
4.4	Kammfiltereffekt .....	20
4.5	Phantomschallquellen .....	21
4.5.1	Pegeldifferenzstereofonie.....	21
4.5.2	Laufzeitstereofonie .....	22
4.6	Kopfbezogenes Koordinatensystem.....	23
4.7	Komplexe Zahlen .....	24
4.8	Digitale Signalverarbeitung.....	25
4.8.1	Vom Zeit- in den Frequenzbereich .....	27
5	Probleme und Fehler, die beim Stereodownmix auftreten.....	30
5.1	Laufzeitunterschiede .....	30
5.2	Veränderung der effektiven Energie .....	32
5.2.1	Erstes Beispiel .....	32
5.2.2	Zweites Beispiel .....	33

5.2.3	Drittes Beispiel .....	34
5.2.4	Ergebnis .....	35
5.3	Differenzpegeländerung .....	36
5.4	Phantomschallquellenverschiebung .....	38
5.4.1	Phantomschallquellen durch Pegeldifferenzen .....	38
5.4.2	Phantomschallquellen durch Laufzeitstereofonie .....	41
6	IRT – Downmix .....	42
6.1	Wesentliche Ansatzpunkte .....	42
6.1.1	Kammfilterkompensation .....	42
6.1.2	Erhaltung der effektiven Energie .....	45
6.1.3	Differenzpegeländerung kohärenter und inkohärenter Signalanteile .....	45
6.1.4	Phantomschallquellenverschiebung .....	45
6.1.5	Eine Lösung für vier Probleme .....	45
6.2	Realisierung des automatischen Downmixes .....	45
6.2.1	Ablauf des Downmixvorgangs .....	46
6.2.2	Der Algorithmus .....	49
6.2.3	Struktur der Implementierung .....	51
7	Analyse und Optimierung des bisherigen Downmixes hinsichtlich der Kammfilterkompensation .....	54
7.1	Subjektiver Eindruck des IRT-Downmix .....	54
7.2	Optimierung durch Analysen im Zeitbereich .....	54
7.2.1	Grafische Darstellung der Bearbeitung im Zeitbereich .....	54
7.2.2	Analyse mit synthetischen Signalen .....	56
7.2.3	Gleichspannungsanteil nach dem Additionsvorgang .....	58
7.2.4	Analyse mit einem stochastischen Signal .....	59
7.2.5	Ein- und Ausschwingvorgang durch die Bearbeitung .....	60
7.3	Optimierung durch Analysen im Spektralbereich .....	62
7.3.1	Grafische Darstellung der Bearbeitung im Spektralbereich .....	62
	Das Testsignal .....	63
	Realisierung der grafischen Darstellung .....	65
7.3.2	Analyse der Bearbeitung der DFT-Koeffizienten .....	65
7.3.3	Optimierung der Kennlinie .....	67
	Theoretische Überlegungen zur Optimierung der Kennlinie .....	67
	Der Algorithmus mit einer Fallunterscheidung .....	70
7.3.4	Analyse der Dämpfung von DFT-Koeffizienten beim Bearbeitungsvorgang .....	71
	Fallanalyse .....	71
	Bearbeitung des größeren DFT-Koeffizienten .....	73
7.3.5	Lösungsansätze .....	76
	Lösung der Dämpfungsproblematik .....	76
	Verschiebung der Position der Fallunterscheidung .....	77
	Korrektur in zwei Durchläufen .....	77
8	Analyse und Optimierung des IRT-Downmixes hinsichtlich der Pegeländerung und Phantomschallquellenverschiebung .....	80

8.1	Analyse .....	80
8.1.1	Pegeländerung .....	81
8.1.2	Phantomschallquellenverschiebung .....	81
8.2	Korrektur.....	82
9	Hörversuch.....	84
9.1	Durchführung.....	84
9.1.1	Versuchsaufbau .....	85
9.1.2	Auswahl des Testmaterials .....	85
	Hörbeispiel 1: Chor .....	85
	Hörbeispiel 2: Klassik1 .....	85
	Hörbeispiel 3: TV-Show .....	85
	Hörbeispiel 4: Hörspiel .....	86
	Hörbeispiel 5: Klassik2 .....	86
	Hörbeispiel 6: Jazz1 .....	86
	Hörbeispiel 7: Jazz2 .....	86
	Hörbeispiel 8: Orgel .....	86
	Hörbeispiel 9: Pop.....	86
	Hörbeispiel A: Intensitätsstereofonie.....	87
	Hörbeispiel B: Laufzeitstereofonie.....	87
	Hörbeispiel C: Kammfilter .....	87
9.1.3	Statistische Auswertung.....	87
9.2	Auswertung der Ergebnisse.....	88
9.2.1	Vergleich IRT-Downmix „5ib“ mit der Downmixversion nach Keil .....	88
9.2.2	Vergleich IRT-Downmix „5ib“ mit dem Downmix nach der ITU-Empfehlung .....	89
9.2.3	Vergleich IRT-Downmix „5j“ mit dem Downmix nach der ITU-Empfehlung .....	89
9.2.4	Vergleich IRT-Downmix „5ib“ mit dem IRT-Downmix „5j“ .....	90
9.2.5	Vergleich IRT-Downmix „5ib“ mit einer von Hand erstellten Stereoverision.....	90
9.2.6	Zusammenfassung der Ergebnisse .....	91
9.3	Nachträgliche Anpassung des Algorithmus .....	92
10	Weitere Überlegungen hinsichtlich eines automatischen Downmixes.....	93
10.1	Lautheit und Dynamik von Stereo- und Mehrkanal-Produktionen .....	93
10.1.1	Zweikanal-Stereo .....	93
10.1.2	Mehrkanal.....	93
10.1.3	Probleme beim automatischen Downmix.....	93
10.2	Verbreitung von Mehrkanal und Stereo .....	94
10.2.1	Ein Radioprogramm im 5.1 Mehrkanal-Format.....	96
11	Anwendungsmöglichkeiten.....	97
11.1	Der IRT-Downmix beim Rundfunk.....	97
11.2	Der IRT-Downmix in der Audioproduktion .....	97
11.3	Der IRT-Downmix beim Rezipienten.....	98
11.4	Der IRT-Downmix als generelles Mischverfahren zweier Signale .....	98
12	Zusammenfassung und Ausblick .....	100

13	Anhang.....	102
13.1	Quelltext .....	103
13.1.1	Quelltext: IRT-Downmix, Version „IRT_5keil“ .....	103
13.1.2	Quelltext: IRT-Downmix, Version „IRT_5ib“ .....	107
13.1.3	Quelltext: IRT-Downmix, Version „IRT_5j“ .....	111
13.1.4	Quelltext: IRT-Downmix, Version „IRT_5r“ .....	115
13.1.5	Direkter Vergleich der Algorithmen.....	119
13.1.6	Quelltext: Funktion zur Darstellung der Bearbeitung der DFT-Koeffizienten.....	121
13.1.7	Quelltext: Ausgabe einer Fallanalyse zweier DFT-Koeffizienten.....	123
13.2	Berechnung der Pegel einer Phantomschallquelle abhängig vom Winkel .....	125
13.3	Blockschaltbild des IRT-Downmix .....	127
	Analyse und Korrektur 1.....	128
	Analyse und Korrektur 2.....	128
	Analyse und Korrektur 3.....	129
	Analyse und Korrektur 4.....	129
13.4	Schriftliche Einführung für den Hörversuch .....	130
	Quellenverzeichnis.....	131
	Abbildungsverzeichnis .....	134
	Glossar .....	137
	Trackliste der DVD-Audio.....	139
	Danksagung .....	140



## 1 Einleitung

Die qualitativen Vorteile des Mehrkanaltons kamen in der Vergangenheit hauptsächlich im Kino zur Geltung. Derzeit hält dieser Surround-Sound zunehmend Einzug in die Haushalte. Längst liegen das Interesse und eine gesteigerte Wertschätzung gegenüber dem vollkommen umhüllenden, auditiven Klangerlebnis nicht mehr nur bei professionellen Tonschaffenden und Hifi-Freaks. Ein Rundgang durch ein Elektronik-Kaufhaus offenbart: der Mehrkanalton ist bereits beim Konsumenten angekommen. Receiver für die Wiedergabe von 7.1 Mehrkanalton, DVDs mit mehrkanaligen Audioinhalten, Video- und Computerspiele mit Surround-Sound sowie Dts- und SACDs sind mittlerweile fester Bestandteil der Ladenregale.

Nicht zuletzt trägt die Verbreitung von HDTV zu einer verstärkten Nachfrage bezüglich einer umfassenderen und besseren Tonwiedergabe bei. Analog zum Wechsel von SDTV zu HDTV, kann also auch eine Erweiterung und Qualitätssteigerung des Tons vollzogen werden.

Die Debatte, ob auch der Rundfunk Inhalte mit Mehrkanalton verbreiten soll, muss ernst genommen werden. Bei der Ausstrahlung mehrkanaligen Tons tritt allerdings ein entscheidendes Problem auf, das den Anlass für diese Arbeit liefert: die Kompatibilität des Mehrkanaltons zum Zweikanal-Stereoton. Denn bei aller Begeisterung, die um das Thema Surround-Sound entstanden ist, bleibt anzunehmen, dass Zweikanal-Stereo auch in Zukunft „der“ Standard bei der Wiedergabe von Audioinhalten bleiben wird. Auch wenn die Mehrkanalwiedergabe bereits erfolgreich in Heimkinos und Luxus-Limousinen zum Einsatz kommt, wird es wenig Anreiz geben, Liefer- und Kleinwagen, Kofferradios oder das Badezimmer zuhause mit mehr als zwei diskret angesteuerten Lautsprechern auszustatten.

Es ist also notwendig, dem Rezipienten neben dem Mehrkanalton auch eine Zweikanalversion für den Stereobetrieb anzubieten. Am einfachsten ist dies durch einen automatischen Stereodownmix zu erreichen. Dieser müsste allerdings ein qualitativ hochwertiges Ergebnis liefern, das objektiv wie subjektiv dem bisher von Hand gemischten Stereoton nicht nachsteht.

Der Anspruch an einen Downmix ist eine Zweikanal-Stereofassung, welche ein sinnvolles Klangabbild des Mehrkanaltons wiedergibt. Dabei ist die akustische Information in dem Bereich zwischen den verbleibenden zwei Lautsprechern so unterzubringen, dass sie beim Hörer möglichst gleich, zumindest ähnlich wahrgenommen wird. Künstlerische Gesichtspunkte wie die Intention der Aufnahme und qualitative Aspekte des Mehrkanaltons, wie die Klangfarbe oder die Lautstärkebalance, sollen so wenig wie möglich beeinflusst werden.

Ziel dieser Arbeit ist es, beim Downmixprozess mit Hilfe der bisherigen Untersuchungen und Entwicklungen, die zu diesem Thema am IRT durchgeführt wurden, ein Ergebnis zu erreichen, das den genannten Ansprüchen gerecht wird und die Kompatibilität des Mehrkanaltons zum Zweikanal-Stereoton sichert.

Diese Ausarbeitung versucht, die teilweise komplexen Vorgänge verständlich darzulegen. Aus diesem Grund ist eine Einführung in die Begriffe und Theorien, auf denen der IRT-Downmix aufbaut, unausweichlich. Diese theoretischen Kapitel am Anfang sind kurz gehalten, praxisnahe Bei-

spiele und Hinweise auf historische Zusammenhänge tragen zur Konkretisierung bei. Leser mit fundierten Kenntnissen in den Bereichen der Tontechnik, Signaladdition und digitalen Signalverarbeitung können die Lektüre dieser Arbeit problemlos mit Kapitel 5 beginnen.

## 2 Mehrkanalton

Es war der 1. Juni 1961, als die beiden Radiostationen WEFM in Chicago und WGFM in Schenectady/NY mit regulären Sendungen in Stereo begannen<sup>1</sup> und dem Hörer damit ein ganz neues Klangerlebnis lieferten. Bis zu diesem Zeitpunkt hatte ein Radiogerät einen Lautsprecher, über den sämtliche Schallinformationen punktförmig wiedergegeben wurden.

Die Zweikanal-Stereofonie stellt aus heutiger Sicht die erste und möglicherweise auch die entscheidendste Etappe des Mehrkanaltons dar. Grund dafür war unter anderem die drastische Verbesserung des Hörerlebnisses bei vergleichsweise geringem Mehraufwand an Technik.

Die ersten Übertragungsversuche mit mehr als zwei Kanälen reichen bis ins 19. Jahrhundert zurück. 1881 übertrug der Ungar Tivadar Puskas in Paris Operaufführungen mit mehreren Telefonleitungen. Die Bell Telephone Company experimentierte 1933/34 bei Konzertaufnahmen mit drei Mikrofonen, deren Signale über drei unabhängige Lautsprecher wiedergegeben wurden und so einen „plastischen“ Klang erzeugten.<sup>2</sup>

In den siebziger Jahren kam die Quadrofonie auf, welche die Begrenzung des akustischen Nutzraums auf den Bereich vor dem Hörer aufheben sollte.

Im Gegensatz zur Stereofonie<sup>3</sup> erreichte die Quadrophonie nie eine große Verbreitung und wird heute oft nur als Vorläufertechnik des späteren Surround-Mehrkanaltons erwähnt.<sup>4</sup> Gescheitert ist diese Technik, neben einem fehlenden bzw. schlechten Trägermedium und einem zu großen finanziellen Aufwand für den Konsumenten, an der unzureichenden Kompatibilität zur Stereofonie.

Das heute gebräuchliche 5.1-Mehrkanalton-Format geht auf die Kinotechnik zurück. Mit der Einführung des 70-mm-Films im Jahr 1970 konnte der Film mit sechs diskreten Audiokanälen auf Magnettonspuren versehen werden. Aus diesen sechs Kanälen entwickelte Dolby das 5.1-Format, aus dem dann die Heimlösungen Dolby Surround<sup>5</sup>, Pro Logic<sup>6</sup> und mit der Einführung der Digitaltechnik DolbyDigital<sup>7</sup> entstanden, die dem Konsumenten den im Kino erprobten Mehrkanalton zugänglich machten.<sup>8</sup>

### 2.1 Tonformate

Die für diese Arbeit relevanten Tonformate sind die Monophonie, die Zweikanal-Stereofonie und der 5.1-Mehrkanalton.

---

<sup>1</sup> vgl. Stereophonic sound 2008

<sup>2</sup> vgl. Große 1981, S. 140

<sup>3</sup> Der Begriff Stereofonie bezieht sich in dieser Arbeit auf die Zweikanal-Stereofonie.

<sup>4</sup> vgl. Quadrophonie 2008

<sup>5</sup> Dolby Surround, auch Dolby Stereo genannt, ist ein von Dolby entwickeltes analoges Mehrkanal-Tonsystem, das durch die Matrixcodierung vier Tonkanäle in einer Stereospur unterbringt. Vgl. Görne 2008, S. 297.

<sup>6</sup> Dolby Pro Logic ist eine Weiterentwicklung von Dolby Surround. Vgl. Görne 2008, S. 298.

<sup>7</sup> Dolby Digital ist ein von Dolby entwickeltes digitales Mehrkanal-Tonsystem, das in der Filmtechnik, auf DVDs und im Rundfunkbereich zum Einsatz kommt. Vgl. Görne 2008, S. 298.

<sup>8</sup> vgl. Birkner 2002, S. 27

### 2.1.1 Monofonie

Das Wort Monofonie kommt ursprünglich aus dem Griechischen und bedeutet übersetzt: die einkanalige Schallübertragung. Diese wurde 1923 mit der ersten Unterhaltungsradiosendung aus der Sendestelle Berlin VOX-Haus eingeführt.<sup>9</sup>

Ein einziger Kanal dient zur Informationsübertragung. Der Schall wird vom Lautsprecher des Empfängers am Wiedergabeort punktförmig abgebildet. Rundfunkempfänger für den Einkanalempfang finden sich auch heute noch im Handel, daher ist die Abwärtskompatibilität zu Mono aller gesendeten Rundfunkformate erforderlich. Bis in die achtziger Jahre hinein wurden von der Musikindustrie noch spezielle Mono-Mischungen der Musiktitel für den Hörfunk erstellt.

Bis heute gibt es Sendeprogramme, die beispielsweise auf Sprachinhalten basieren, wie der Nachrichtensender des Bayerischen Rundfunks „Bayern 5“, die ausschließlich im Monoformat ausgestrahlt werden.

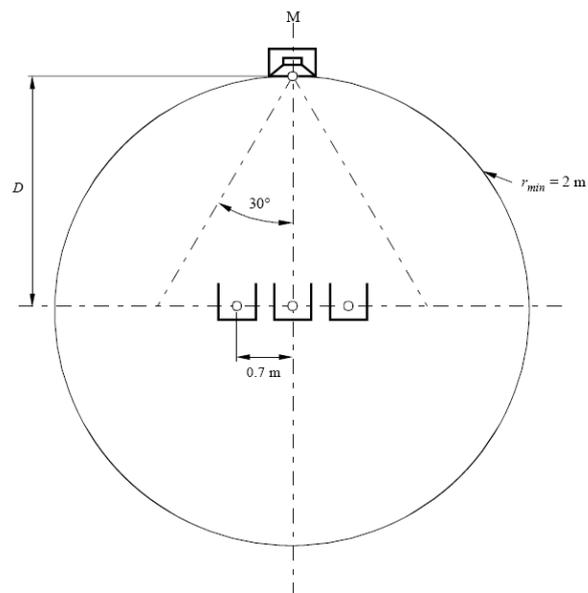


Abbildung 2.1-1 Empfehlung der Lautsprecheraufstellung und Abhörposition für Mono nach ITU-R BS.1116

### 2.1.2 Zweikanal-Stereofonie

Das Wort Stereofonie ist vom griechischen Wort stereos abgeleitet, das mit „räumlich“ und „ausgedehnt“ übersetzt werden kann. Es steht für eine Technik, mit deren Hilfe akustische Signale durch die Wiedergabe über zwei Schallquellen räumlich abgebildet werden.

Nachdem es schon im 19. Jahrhundert Versuche mit mehreren Audiokanälen gegeben hatte, war es Alan Dower Blumlein, der 1931 die Stereoschallplatte zum Patent anmeldete.<sup>10</sup>

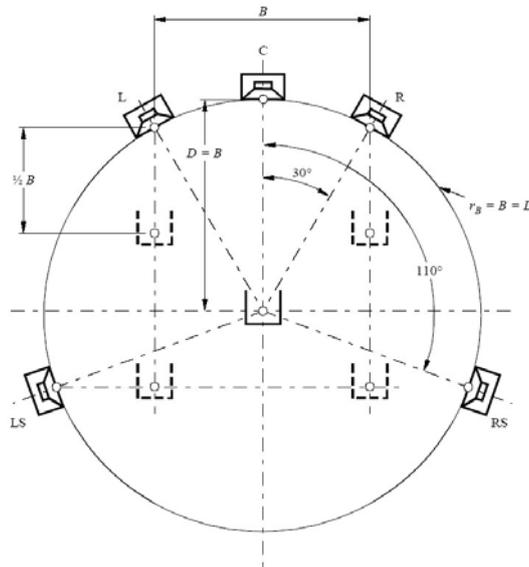
Entscheidend war dabei die Technik, wie Blumlein die beiden Kanäle trennt und zugleich eine einfache Monokompatibilität erreicht. Das sogenannte MS-Verfahren stellt die beiden Signale für den linken und den rechten Lautsprecher (L und R) aus einem Mittensignal und einem Seitensignal her (M und S). Das Seitensignal definiert sich über die Differenz der beiden Signale L und R, das Mittensignal über deren Summe.

<sup>9</sup> vgl. Birkner 2002, S. 22

<sup>10</sup> vgl. Birkner 2002, S. 23



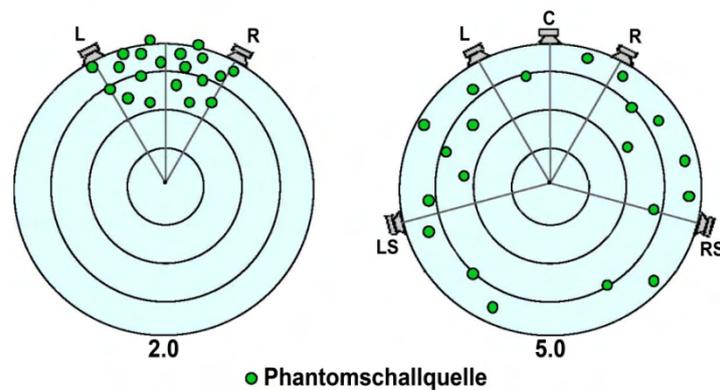
Die Aufstellung der Lautsprecher baut auf der 3/2-Empfehlung der ITU auf. Diese sieht neben dem linken und rechten Kanal (L und R) einen Center-Kanal (C) im vorderen Halbkreis vor. Rückwärtige Schallereignisse werden über zwei Surroundkanäle (LS und RS) realisiert.<sup>12</sup> In Abbildung 2.1-3 ist die Aufstellung der Lautsprecher sowie die Abhörposition grafisch dargestellt.



**Abbildung 2.1-3** Empfehlung für die Lautsprecheranordnung und Abhörposition von 3/2-Mehrkanalton nach ITU-R BS.775-2

Neben den 5 diskreten Kanälen aus der 3/2 Empfehlung, bietet der 5.1-Standard die Möglichkeit zur Erweiterung um einen Tiefbasskanal (engl.: Low Frequency Effect, LFE), der optional eingesetzt werden kann.

Das 5.1-Mehrkanalformat erweitert die Zweikanal-Stereofonie nicht wie oft behauptet um eine zusätzliche akustische Dimension, vielmehr werden die beiden, schon in der Zweikanal-Stereofonie vorhandenen Dimensionen erst vollständig genutzt. Schallereignisse können beim 5.1-Mehrkanalton - je nach Ort mehr oder weniger gut - in der kompletten Horizontalebene platziert werden.<sup>13</sup>



**Abbildung 2.1-4** Möglichkeiten der Anordnung von Phantomschallquellen in 2.0-Stereo oder 5.0-Mehrkanal.

<sup>12</sup> vgl. ITU-R BS.775-2

<sup>13</sup> vgl. Theile 2004, S. 13 f.

Die Nutzung einer weiteren Dimension, zum Beispiel der Höhe, wird erst durch noch aufwendigere Mehrkanaltonsysteme möglich, wie beispielsweise die Aurofonie. Hier werden über den fünf Lautsprechern in Kopfhöhe (L, C, R, LS, RS) weitere, bis zu fünf Lautsprecher angebracht.

## 2.2 Mehrkanal beim Rundfunk

Am 1. Januar 2003 strahlte der ORF als erste öffentlich rechtliche Rundfunkanstalt Europas das Neujahrskonzert im 5.1-Mehrkanalton über DVB aus.<sup>14</sup>

Im Rahmen eines Pilotprojekts auf der IFA 2003 präsentierten der Bayerische und der Westdeutsche Rundfunk in Zusammenarbeit mit dem IRT erstmals 5.1-Produktionen im Radioservice über DVB-S.<sup>15</sup>

Das 5.1-Mehrkanalformat hat in den letzten Jahren also Einzug in die Rundfunkanstalten gehalten. Während im Fernsehen immer mehr Sendungen mit dem mehrkanaligen Ton ausgestattet werden, ist das Angebot im Bereich Hörfunk allerdings immer noch eher spärlich.

Vor allem produktionsseitig erfordert die Übertragung von 5.1-Mehrkanalton eine Umstellung. Das Sendematerial muss besorgt oder selbst produziert werden, technische Geräte müssen ausgetauscht oder für die Verarbeitung von Mehrkanalton angepasst werden, Tonträger mit genügend diskreten Kanälen werden benötigt und die Übertragung des mehrkanaligen Sendeprogramms zum Rezipienten muss realisiert werden.

### 2.2.1 Produktion von Mehrkanalton

Ein Großteil der Studios für künstlerische Produktionen und der Ü-Wägen bei den öffentlich-rechtlichen Rundfunkanstalten in Deutschland sind mittlerweile technisch für Mehrkanaltonproduktionen gerüstet.

Vor allem im Bereich von Hörspielen und der ersten Musikproduktion werden diese technischen Möglichkeiten auch - je nach Rundfunkanstalt unterschiedlich stark - benutzt. Der Westdeutsche Rundfunk ist dafür bekannt, einen Großteil der Hörspiele in 5.1-Mehrkanalton zu produzieren. Der Südwestrundfunk produziert hingegen nach einigen Testprojekten in 5.1-Mehrkanalton die meisten Hörspiele wieder in Stereo. Es sind wirtschaftliche Interessen und ein begrenztes Budget, die hier Grenzen stecken. Nur wenn eine Vermarktung über den Sendebetrieb hinaus angedacht ist, wird die Mehrkanaltonproduktion bevorzugt.

### 2.2.2 Tonträger

Eine Voraussetzung für die Übertragung von Mehrkanalton ist die Verfügbarkeit von Tonträger, die eine entsprechende Anzahl von diskreten Kanalspuren haben. Die DVD bietet die Möglichkeit, digitale Tonsignale als lineare PCM<sup>16</sup> oder datenreduziert im DolbyDigital- und DTS<sup>17</sup>-Format

---

<sup>14</sup> vgl. Stoll 2005, S. 50

<sup>15</sup> vgl. Kapitel 2.2.3 Möglichkeiten der Übertragung

<sup>16</sup> Die „Pulse Code Modulation“ (PCM) ist eine verlustfreie Modulationsform, bei der ein Signal binär kodiert wird. Das Verfahren wird für Audio- und Videosignale verwendet. Vgl. Dickreiter, Band 2 1997, S. 271.

<sup>17</sup> DTS bezeichnet das von „Digital Theater Systems“ entwickelte, digitale Mehrkanaltonsystem. Dieses kommt vorrangig im Kino, auf CDs und DVDs zum Einsatz. Vgl. Görne 2008, S. 298.

abzulegen. Auch die CD kann mit einem DTS-Bitstream mit 1,44 Mbit/s Mehrkanalton speichern. Auf der speziell für den Mehrkanalton entwickelten SACD liegt das Audiomaterial im Direct Stream Digital (DSD)<sup>18</sup>-Format vor.<sup>19</sup>

Bei Rundfunkproduktionen werden Speichermedien mit 8-Spur Format bevorzugt, wie zum Beispiel das Digital Tape Recording System (DTRS)<sup>20</sup> von Tascam oder die Dolby E-Codierung<sup>21</sup>, die acht diskrete Kanäle auf einem digitalen Stereokanal unterbringt. Die Spurbelegung der acht Kanäle ist festgelegt. Die Kanäle 1 bis 6 sind für den 5.1-Mehrkanalton reserviert, die Kanäle 7 und 8 werden in der Regel für eine Stereo-Mischung verwendet.<sup>22</sup>

Neben den genannten Datenträgern spielt der Audio-Massenspeicher (AMS), der die Daten in Form einer digitalen Datei auf einer Festplatte zum Beispiel auf einem Server speichert, eine immer bedeutsamere Rolle.

### 2.2.3 Möglichkeiten der Übertragung

Für die digitale Verbreitung auditiver Inhalte bieten sich derzeit drei Verfahren an: DAB<sup>23</sup> (Digital Audio Broadcasting), DVB<sup>24</sup> (Digital Video Broadcasting) sowie die Stream-Technologie über das Internet.

Das Institut für Rundfunktechnik (IRT) hat bereits 1997 und 2004 DAB erfolgreich hinsichtlich seiner Mehrkanalfähigkeit untersucht. Derzeit stehen aber noch keine entsprechenden Empfänger zur Verfügung. Auch hat sich DAB in der Vergangenheit in Deutschland schwer getan, sich gegen den UKW-Empfang durchzusetzen. Jedoch wurde auf der internationalen Wellenkonferenz RRC 06 im Juni/Juli 2006 dem Wunsch der Bundesrepublik Deutschland entsprochen, zukünftig zwei weitere nationale Bedeckungen mit rund 12 - 18 DAB- oder entsprechend vielen DMB<sup>25</sup>-Programmen zu ermöglichen. Durch die Verlegung von DVB-T aus dem VHF- in das für DVB-T besser geeignete UHF-Band entstehen weitere sieben Bedeckungen mit denen 42 - 63 DAB-Prog-

---

<sup>18</sup> Bei Direct Stream Digital Aufnahmen kommen so genannte Delta-Sigma-Verfahren zum Einsatz. Delta-Sigma-Analog-Digital-Wandler sind bei gleicher Qualität einfacher und kostengünstiger zu bauen als konventionelle A/D-Wandler bzw. erreichen bei gleichem finanziellem Aufwand eine deutlich bessere Qualität. Die Wortbreite des digitalisierten Audiosignals beträgt hierbei lediglich 1 Bit, bei einer Abtastrate von 2,8224 MHz (64\*44,1kHz). Vgl. Super Audio Compact Disc 2008.

<sup>19</sup> vgl. Sandner 2003, S. 2 f.

<sup>20</sup> Digital Tape Recording System (DTRS) wurde 1993 von Tascam eingeführt. Es handelt sich hierbei um ein Aufnahmesystem, welches es ermöglicht auf einer Hi8-Kassette acht Tonspuren digital aufzuzeichnen.

<sup>21</sup> Dolby E ist ein von Dolby entwickeltes, auf den Sendebetrieb ausgerichtete, mehrkanaliges Codierformat. Dabei werden acht Kanäle (20 Bit, 48 kHz) auf einem digitalen Stereokanal untergebracht, so dass mit einem AES/EBU-Kanal acht diskrete Kanäle übertragen werden können. Vgl. Birkner 2002, S. 144.

<sup>22</sup> vgl. Sandner 2003, S. 3

<sup>23</sup> Digital Audio Broadcasting (DAB) ist ein digitaler Übertragungsstandard für die Ausstrahlung von Audioinhalten. Vgl. Digital Audio Broadcasting 2008.

<sup>24</sup> Digital Video Broadcasting (DVB) ist ein digitaler Übertragungsstandard für die Verbreitung von Video- und Audioinhalten. Vgl. Digital Video Broadcasting 2008.

<sup>25</sup> Digital Multimedia Broadcasting (DMB) ist ein von der Robert Bosch GmbH und dem Heinrich-Hertz-Institut entwickeltes, digitales Übertragungssystem, das, basierend auf dem Hörfunkstandard DAB, sowohl für die terrestrische Übertragung (T-DMB) als auch die Satellitenübertragung (S-DMB) verwendet werden kann. Gegenüber DAB ist die Übertragung audiovisueller Inhalte möglich.

ramme verbreitet werden können. Ob die Umsetzung schnell erfolgt ist fraglich. Der Verbraucherzentrale Bundesverband (vzvb) und der ADAC raten bereits jedoch bei Neuanschaffungen nur noch Kombigeräte zu kaufen, die neben UKW auch DAB empfangen können.<sup>26</sup>

Auch DVB ist nach einer Erweiterung des DVB-Standards (ETSI, ETR 154) durch die optionale 5.1-Mehrkanalton-Übertragung im Dolby-Digital-Format seit 1997 mehrkanalfähig.<sup>27</sup> Es gibt mehrere spezifische DVB-Varianten für unterschiedliche Übertragungswege, die sich hauptsächlich im Modulationsverfahren und der Fehlerkorrektur unterscheiden. Für die Übertragung von Mehrkanalton eignen sich DVB-S<sup>28</sup>, DVB-C<sup>29</sup> und DVB-T<sup>30 31</sup>.

Die DVB-T zugewiesenen Frequenzen sind in Deutschland auf Grund ihrer noch relativ geringen Anzahl den Fernsehsendern vorbehalten. Radiodienste sind nicht geplant, auch wenn dies theoretisch möglich wäre. In den Niederlanden, Belgien und Norwegen werden Radioprogramme auch über DVB-T ausgestrahlt.

Über Satellit und Kabel sind hingegen auch in Deutschland digitale Radioprogramme via DVB empfangbar. Eine beträchtliche Anzahl zur Verfügung stehender, mehrkanalfähiger Set-Top-Boxen<sup>32</sup>, die eine große Reichweite garantieren, macht die Übertragung für Rundfunkanstalten interessant. Ein Nachteil für den Rezipienten ist aber, dass diese für den Fernsehbetrieb ausgelegten Empfangsgeräte hinsichtlich der Bedienung für Hörfunkzwecke nicht optimal geeignet sind. Eine Navigation ohne Monitor ist bei den meisten Set-Top-Boxen schwierig.

Ein neuer Satellitentransponder für ARD-Hörfunkprogramme wurde Anfang September 2005 in Betrieb genommen. Der Bandbreitengewinn ermöglicht neben dem MPEG-1 Layer 2-Datenstrom (Stereo) auch einen Dolby-Digital-Datenstrom mit 448 kbit/s (Stereo und Mehrkanalton).

### 2.3 Notwendigkeit des Downmixes

Wenn Radioprogramme wie auch Fernsehprogramme mit Mehrkanalton ausgestrahlt werden, wird man selbst auf längere Sicht nicht auf eine Stereo-Version des Audioinhalts verzichten können. Verschiedene Lebensbereiche, in denen eine solche mediale Unterhaltung durchaus in Anspruch genommen wird, wie zum Beispiel die Küche oder das Badezimmer, sind für eine Mehrkanaltonwiedergabe nicht geeignet. Es darf demzufolge bezweifelt werden, ob der Mehrkanalton jemals eine so dominante Rolle wie die Stereophonie einnehmen wird. Daher ist es unabdingbar dem Hörer neben dem Service „Mehrkanalton“ auch immer eine monokompatible Stereomischung zur Verfügung zu stellen.

Diese Bedingung setzt eine Kompatibilität des Mehrkanaltons mit der Stereowiedergabe voraus. Die Kompatibilität kann entweder empfängerseitig durch eine Matrizierung der Kanäle des Mehr-

---

<sup>26</sup> vgl. Digital Audio Broadcasting 2008

<sup>27</sup> vgl. Stoll 2005, S. 50

<sup>28</sup> DVB-S ist für die Übertragung durch direktstrahlende Satelliten ausgelegt.

<sup>29</sup> DVB-C ist für die Übertragung über Kabelnetze ausgelegt.

<sup>30</sup> DVB-T ist für die Übertragung durch terrestrische Senderketten im VHF- bzw. UHF-Bereich ausgelegt.

<sup>31</sup> vgl. Schmidt 2003 S. 203

<sup>32</sup> Ein Gerät (auch Beistellgerät genannt), das für den Betrieb eines anderen Gerätes - meist ein Fernseher - erweiterte Nutzungsmöglichkeiten bietet.

kanaltons auf zwei Kanäle hergestellt werden, oder senderseitig durch die Übertragung sowohl einer Mehrkanalversion als auch einer Stereoverversion.

Wie in den folgenden Kapiteln zu sehen sein wird, bringt die Matrizierung des Mehrkanaltons auf zwei Kanäle Probleme mit sich. Bei einer senderseitigen Produktion können diese vermieden und die Stereoverversion vor der Übertragung kontrolliert werden. Angesichts der großen Zahl der Hörer, die auch in Zukunft Rundfunkangebote in Stereo nutzen werden, stellt dies einen großen Vorteil dar.

Senderseitig kann eine Stereoverversion entweder durch eine separate Mischung des Audiomaterials oder durch einen automatischen Downmix erstellt werden. Von künstlerischer Seite ist einer separaten Mischung klar der Vortritt zu gewähren. Im Falle einer Live-Sendung hieße dies aber: zwei Ü-Wägen und mehr Personal, das an der Mischung beteiligt ist. Leider ist dies aus technischen oder logistischen Gründen in manchen Fällen, aus wirtschaftlichen Gründen in den meisten Fällen, nicht möglich.

Alternativ wird nur eine Mehrkanalmischung erstellt und mit einem automatischen Downmix die dazugehörige Stereoverversion erzeugt. Würde ein solcher Downmix in Echtzeit errechnet, könnte dieser auch gleich vom mischenden Tonmeister oder Toningenieur kontrolliert werden.

Dieses Verfahren ist aber angesichts der großen Stereohörerschaft nur zu verantworten, würde der Downmix ein allgemein anerkanntes und subjektiv wie objektiv gutes Ergebnis liefern.

## 3 Downmix

Wie bereits in Kapitel 2.3 angedeutet, versucht ein Downmixverfahren die einzelnen diskreten Kanäle des Mehrkanaltons mit sinnvollen Anteilen auf zwei Kanäle zu matrizieren. Dabei soll möglichst das akustische Abbild der vorderen Lautsprecher nicht zerstört und die Schallinformation der hinteren Kanäle geschickt in das vordere Klangbild miteingebaut werden. Verschiedene Matrizierungsparameter definieren die Anteile der einzelnen Kanäle bei der Matrizierung.

Neben der Möglichkeit die Stereoersion manuell von Hand zu mischen, wird zwischen zwei automatischen Downmixverfahren unterschieden:

1. Passive Verfahren, die den Mehrkanalton ohne Signalanalyse und mit statischen Parametern zu zwei Kanälen zusammenfassen.
2. Aktive Verfahren, die sich dynamisch an das Ausgangsmaterial anpassen können, um ein bestmögliches Ergebnis zu erreichen.

### 3.1 Handmix

Der Handmix ist eine individuelle Zweikanalmischung, die unabhängig von der Mehrkanalmischung aus den einzelnen ungemischten Quellsignalen erstellt wird. Dafür kann auch eine andere, für die Stereomischung geeignete Anordnung des Hauptmikrofons verwendet werden. Die Signale der Stützmikrofone können unabhängig bearbeitet und hinzu gemischt werden.

Diese Variante des Downmix erlaubt ein Höchstmaß an künstlerischer Entfaltung des Tonmeisters oder –ingenieurs, so dass sich eine Handmixfassung teilweise auch stark von der Mehrkanalmischung unterscheiden kann. Probleme, die durch eine Matrizierung von mehreren Audiosignalen entstehen, können teilweise von vornherein vermieden werden.

### 3.2 Automatische passive Downmixverfahren

Passive Downmixverfahren zeichnen sich durch feste Einstellungen und Parameter aus. Sie arbeiten also unabhängig von der Beschaffenheit des Ausgangsmaterials.

#### 3.2.1 ITU Downmix

In der Veröffentlichung „BS.775“ der ITU wird neben einer Empfehlung zur Lautsprecherpositionierung und der optimalen Abhörposition (engl.: Sweet Spot) auch ein passives Downmixverfahren vorgeschlagen, das die Abwärtskompatibilität sicherstellen soll.

Dabei werden dem linken Kanal der Centerkanal und der linke Surroundkanal jeweils mit einer Absenkung von -3dB zugemischt. Der rechte Kanal wird auf die gleiche Weise matriziert. Die Abbildung 3.2-1 zeigt das zugehörige Blockschaltbild.

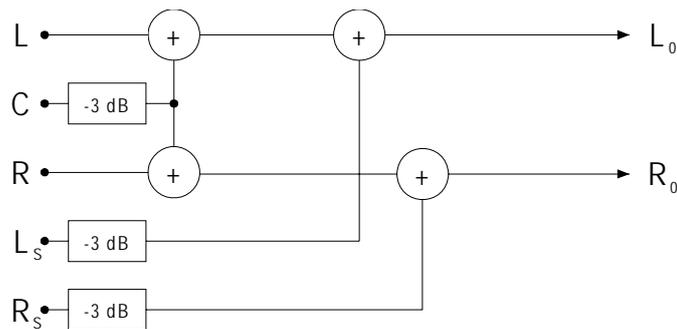


Abbildung 3.2-1: Blockschaltbild des Downmix nach ITU BS.775

Wie man dem Blockschaltbild entnehmen kann, findet der LFE-Kanal bei der Matrizierung keine Berücksichtigung. Dies ist auch nicht nötig, da der LFE-Kanal eindeutig als nicht zwingend notwendiger, zusätzlicher Effektkanal definiert ist.<sup>33</sup>

Alle Angaben in den ITU-Recommendations sind Empfehlungen. So auch die Parameterwertangaben von -3dB, die der Anwender individuell den Anforderungen anpassen kann.

Der Österreichische Rundfunk ORF führte umfangreiche Hörtests mit unterschiedlichem Testmaterial durch und modifizierte die Parameter für Klassik- und Radiokunstproduktionen wie folgt:

Klassikproduktionen

$$L_0 = L + 0,75 \cdot C + 0,63 \cdot L_s$$

$$R_0 = R + 0,75 \cdot C + 0,63 \cdot R_s$$

Radiokunstproduktion

$$L_0 = L + 0,79 \cdot C + 0,5 \cdot L_s$$

$$R_0 = R + 0,79 \cdot C + 0,5 \cdot R_s$$

### 3.2.2 Dolby Pro Logic II

Dolby Pro Logic II ist eine erweiterte Form von Pro Logic, die im Jahr 2000 von Dolby vorgestellt wurde. Sie erweitert Pro Logic um einen zweiten Surroundkanal, so dass ein 5.0 Mehrkanalton encodiert werden kann. Dolby bezweckt mit seinem Encoder nicht nur einen Downmix auf die zwei Kanäle  $L_t$  (left total) und  $R_t$  (right total), sondern ermöglicht auch die Wiederherstellung des Mehrkanaltons aus dem entstandenen Stereoton  $L_t$  und  $R_t$ .

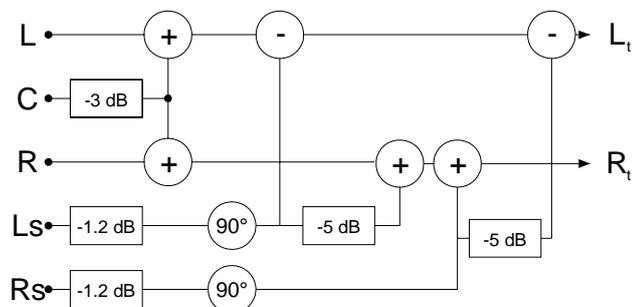


Abbildung 3.2-2 Blockschaltbild des Dolby Pro Logic II Encoders

<sup>33</sup> vgl. ITU-R BS.775-2

Um  $L_t$  und  $R_t$  zu erhalten, wird das Centersignal um 3 dB abgesenkt und dem linken wie dem rechten Signal zugemischt. Die Surroundsignale LS und RS werden jeweils um 1,2 dB abgesenkt und um  $90^\circ$  in der Phase gedreht. LS wird dann sowohl vom linken Signal subtrahiert als auch um 5 dB abgesenkt dem rechten Signal zugemischt. Der Matrizevorgang beim rechten Signal verläuft auf die gleiche Weise.

Die Surroundanteile in  $L_t$  und  $R_t$  sind durch die Subtraktion beim linken und der Addition beim rechten Kanal um  $180^\circ$  zueinander phasenverschoben. Dieser frequenzabhängige Zeitversatz sorgt dafür, beim Decodierprozess eine Kanaltrennung durchführen zu können.

### 3.3 Automatische aktive Downmixverfahren

Aktive Downmixverfahren zeichnen sich durch eine dynamische Anpassung der Parameter abhängig vom Ausgangssignal aus.

#### 3.3.1 Lexicon Logic7

Das von David Griesinger entwickelte Verfahren Logic7 der Firma Lexicon beinhaltet neben einem Downmix auch die Möglichkeit zum Upmix. Dabei arbeitet der Algorithmus sehr flexibel. Der Mehrkanalton kann sowohl auf ein Mono- als auch auf ein Stereosignal herunter gemischt werden. Aus einem Stereodownmix können dann beispielsweise wieder bis zu acht Kanäle decodiert werden.

Griesinger sind für seinen Downmix besonders die Erhaltung der effektiven Energie, eine möglichst geringe Verfälschung der Phantomschallquellen sowie die dynamische Anpassung der Dämpfungsparameter von den Surroundkanälen und dem Centerkanal wichtig.

Diese Anforderungen realisiert Griesinger mit einem aktiven Downmixalgorithmus, der über die variablen Koeffizienten  $m_r$  und  $m_l$  den Anteil des Centerkanals im Downmix regelt und mit den Koeffizienten  $m_i$  und  $m_s$  den Anteil der Surroundkanäle anpasst. Die Surroundkanäle werden für den Downmix gemischt. Für den linken Kanal wird ein Anteil von 0,91 des linken Surroundkanals mit einem Anteil von -0,38 des rechten Surroundkanals addiert. Die Mischung für die rechte Seite erfolgt analog dazu. Mit dieser Vorgehensweise bleiben die Pegel der hinteren beiden Kanäle im Downmix erhalten. Durch die Phasendrehung von  $90^\circ$  wird wie bei Dolby's Pro Logic II die spätere Trennung der hinteren und vorderen Kanäle möglich.<sup>34</sup> Den kompletten Aufbau des Algorithmus zeigt die Abbildung 3.3-1.

---

<sup>34</sup> vgl. Griesinger 2003, S. 14

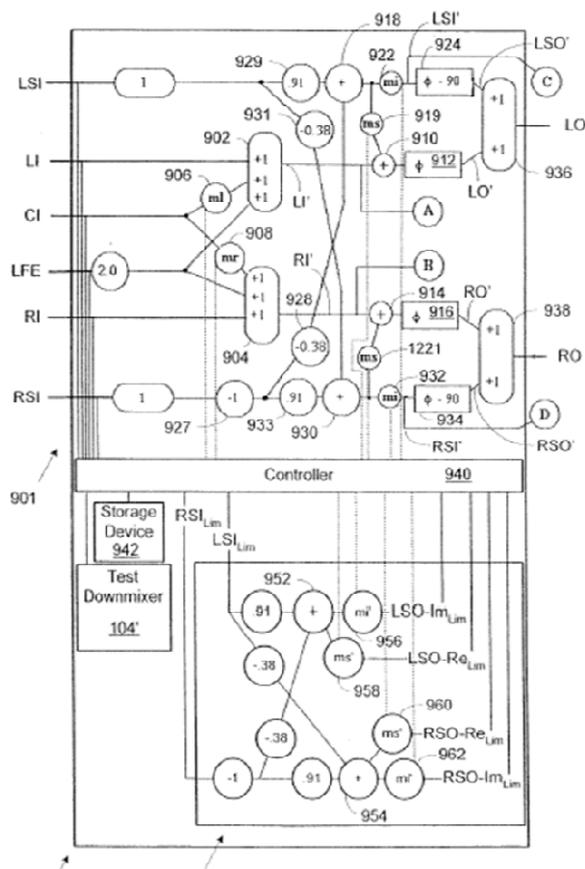


Abbildung 3.3-1 Blockschaltbild Logic7

### 3.3.2 Coding Technologies

Auch die Firma Coding Technologies, die im Sektor der Audiocodec-Entwicklung und -Optimierung tätig ist, hat die Entwicklung eines eigenen Downmixalgorithmus begonnen.

Ausgehend von den Parametern des ITU-Downmixes (Abbildung 3.2-1), wird nach der Addition der Energiegehalt in 28 Frequenzbändern<sup>35</sup> analysiert und mit dem Energiegehalt der einzelnen Kanäle des Ausgangsmaterials verglichen. Auf diese Weise können Anhebungen und Absenkungen ermittelt und durch Amplitudenkorrektur in dem betreffenden Teilband kompensiert werden. Eine Klangfarbenveränderung durch einen Kammfiltereffekt kann so begrenzt werden. Die Korrektur erfolgt aber nur zu einem sinnvollen Anteil.<sup>36</sup> Ein sich vollständig auslöschendes Signal würde sonst einen unendlich großen Korrekturfaktor hervorrufen.

Leider hat die Firma Coding Technologies nach der Übernahme von Dolby das Projekt vorläufig eingestellt.

<sup>35</sup> Die 28 Frequenzbänder sind von der Bark-Skala (vgl. Zwicker & Feldtkeller 1967) abgeleitet, die dem menschlichen Hörverhalten angepasst ist. Die Auflösung des menschlichen Gehörs beträgt im hörbaren Frequenzbereich 27 Bänder. Mit 28 Bändern wird das menschliche Auflösungsvermögen übertroffen.

<sup>36</sup> vgl. Deigtmöller 2006, S. 27

### 3.4 Ergebnisse vorangegangener Untersuchungen am IRT

Im Vorfeld dieser Diplomarbeit wurden bereits zwei Arbeiten zum Thema Downmix verfasst. Eine von Jörg Deigmöller mit dem Thema „Subjektive und objektive Untersuchungen zum 2.0-Downmix von 5.1-Mehrkanal-Produktionen“<sup>37</sup> und die zweite von Martin Keil, der mit dem Thema „Realisierung eines Stereo-Downmixes aus einem 5.1 Signal für Anwendungen im Rundfunk“<sup>38</sup> einen Algorithmus entwickelt hat, welcher Basis dieser Arbeit ist. Sowohl Deigmöller als auch Keil haben im Rahmen ihrer Diplomarbeit einen Hörversuch durchgeführt, der die subjektive Qualität derzeit am Markt befindlicher Downmixalgorithmen evaluieren sollte. Auch der Vergleich mit einer von Hand erstellten Stereo-Mischung wurde nicht gescheut. Im Folgenden sollen die Ergebnisse dieser Untersuchungen kurz erörtert werden.

#### 3.4.1 Erster Hörversuch zum Thema Downmix am IRT

Jörg Deigmöller hatte zum Ziel, mit seinem Hörversuch subjektive Qualitätsunterschiede zwischen den bekannten Downmixverfahren Dolby Pro Logic II, Lexicon Logic7, dem Verfahren von Coding Technologies und der Downmixempfehlung der ITU aufzudecken.

Für die Bewertung wurde ein Paarvergleich (A-B-Vergleich) gewählt, da weder das 5.1-Ausgangsmaterial noch eine der Downmixversionen als Referenz angesehen werden kann. Die Testpersonen konnten die Qualität anhand des Verhältnisses der beiden im Paarvergleich dargebotenen Zweikanalversionen subjektiv bewerten. Hierfür wurde eine von der ITU empfohlene siebenstufige Skala von -3 für „viel schlechter“ bis +3 für „viel besser“ verwendet.<sup>39</sup> Zusätzlich zur Bewertung standen fünf Attribute zur Auswahl, von denen bis zu zwei angekreuzt werden konnten, die für die Entscheidungsfindung ausschlaggebend waren.

Um im Rahmen dieser Untersuchung ein möglichst breites Spektrum abzudecken, wurden acht verschiedene Hörbeispiele aus den Genres Klassik, Experimental, Jazz, Alternative und Radiokunstproduktion ausgewählt.

#### Ergebnis des Hörversuchs

Mit Ausnahme von Dolby Pro Logic II fielen die Unterschiede der Downmixverfahren insgesamt gering aus. Das Verfahren von Dolby ging klar als schlechtestes Downmix aus dem Hörversuch hervor. Logic7, das Verfahren der ITU und der Handmix erzielten vergleichbare Ergebnisse.

Coding Technologies schnitt hingegen etwas besser ab als die Konkurrenz. Diese Ergebnisse zeigen, dass aktive Downmixalgorithmen den passiven überlegen sind. Allerdings stellen Algorithmen, die einen späteren Upmix durch eine Phasenveränderung beim Downmix ermöglichen, eine Einschränkung für die erstellte Zweikanalversion dar.

Als das entscheidendste Kriterium bei der Bewertung wurde das Attribut Klangfarbe angegeben.<sup>40</sup>

Dass der Handmix nicht besser bewertet wurde als automatische Downmixverfahren, ist das eigentlich sensationelle an dieser Untersuchung. Genreabhängig wurde der Handmix sogar schlechter bewertet. Das darf jedoch keines Falls darüber hinweg täuschen, dass nur der Hand-

---

<sup>37</sup> Deigmöller 2006

<sup>38</sup> Keil 2007

<sup>39</sup> vgl. ITU-R BS.1284 1997

<sup>40</sup> vgl. Deigmöller 2006, S. 126 ff.

mix ein Höchstmaß an künstlerischer Individualität besitzen kann und daher, wenn er zur Verfügung steht, bevorzugt werden sollte. Dennoch sind aktuelle Downmixverfahren hinsichtlich ihrer technischen Qualität in der Lage mit einem Handmix zu konkurrieren und stellen deshalb eine sinnvolle Alternative dar, falls die Mischung der Zweikanalversion per Hand nicht möglich ist.

### 3.4.2 Zweiter Hörversuch zum Thema Downmix am IRT

Ziel des Hörversuchs von Keil war die Verifikation des im Rahmen seiner Diplomarbeit entwickelten Downmixalgorithmus. Hierfür wurde ein fast identisches Verfahren gewählt wie bei dem Versuch von Deigmöller. Im Gegensatz zur umfangreichen Testdurchführung Deigmöllers, wo alle Downmixverfahren miteinander verglichen wurden, erfolgte der Paarvergleich nur zwischen dem neu entwickelten Algorithmus und den drei bekannten Verfahren Dolby Pro Logic II, Coding Technologies und der Empfehlung der ITU.

Auf Grund der geringen Teilnehmerzahl von acht Testpersonen kann das Ergebnis lediglich als Tendenz gewertet werden.

Im Vergleich zum ITU-Downmix wurde der neue Algorithmus je nach Genres deutlich besser bewertet. Vor allem in den Bereichen der Klassik und der Radiokunstproduktion konnte Keils Downmix gegenüber der ITU-Empfehlung überzeugen. Auch gegenüber Dolby Pro Logic II schnitt das neue Verfahren von Keil eindeutig besser ab. Beim Vergleich mit dem aktiven Downmixverfahren von Coding Technologies konnten die Testpersonen kaum einen qualitativen Unterschied zum neuen Downmix ausmachen.<sup>41</sup>

Zusammenfassend lässt sich sagen, dass das Verfahren von Keil mit seiner dynamischen Signalverarbeitung vor allem gegenüber den beiden passiven Downmixverfahren klar vorne liegt. Dieses gute Ergebnis ist Grund für die vorliegende Arbeit, die den neuen Algorithmus weiter optimieren soll.

---

<sup>41</sup> vgl. Keil 2007, S. 52 ff.

## 4 Technische Grundlagen und Begriffe

Zum Verständnis der vorliegenden Arbeit sind verschiedene technische Grundlagen und Begriffe notwendig, die zu erläutern sind.

### 4.1 Korrelation

Die Korrelation ist ein Maß für die Ähnlichkeit zweier Signale  $a(t)$  und  $b(t)$  im zeitlichen Verlauf. Sie beschreibt wie gleich oder ungleich sich diese zwei Signale sind. Pegelunterschiede der beiden Signale spielen dabei keine Rolle.<sup>42</sup> Der normierte Korrelationsfaktor oder Korrelationskoeffizient  $\Phi_{a,b}(\tau)$  ist dimensionslos und berechnet sich mit Hilfe der normierten Kreuzkorrelation.<sup>43</sup> Der Wert von  $\tau$  beschreibt dabei die zeitliche Verschiebung des Signals  $b(t)$  gegenüber  $a(t)$ .

$$\Phi_{a,b}(\tau) = \lim_{T \rightarrow \infty} \frac{\frac{1}{2T} + \int_{-T}^{+T} a(t) \cdot b(t + \tau) dt}{\frac{1}{2T} + \sqrt{\int_{-T}^{+T} a^2(t) dt \cdot \int_{-T}^{+T} b^2(t) dt}} \quad (4.1)$$

Der Wertebereich des normierten Korrelationskoeffizienten  $\Phi_{a,b}(\tau)$  liegt zwischen 1 und -1, wobei der Wert  $\Phi_{a,b}(\tau) = 1$  eine lineare Abhängigkeit der beiden Signale angibt, das heißt die beiden Signale  $a(t)$  und  $b(t)$  haben eine identische Kurvenform. Der Wert  $\Phi_{a,b}(\tau) = 0$  zeigt an, dass keinerlei lineare Abhängigkeit vorliegt. Die beiden Signale  $a(t)$  und  $b(t)$  sind dann unabhängig voneinander – also orthogonal zueinander. Weist der Korrelationskoeffizient den Wert  $\Phi_{a,b}(\tau) = -1$  auf, so liegt eine gegenphasige lineare Abhängigkeit vor, das bedeutet die beiden Signale  $a(t)$  und  $b(t)$  sind identisch, aber um  $180^\circ$  phasenverschoben. Liegt der Wert zwischen -1 und 0 bzw. zwischen 0 und 1, so bestehen die beiden Signale  $a(t)$  und  $b(t)$  sowohl aus voneinander abhängigen als auch voneinander unabhängigen Anteilen.

Korrelationsmessgeräte untersuchen die Signale nur mit einer sehr kleinen Integrationszeit von  $T \leq 1s$  und können auch nur für diesen Zeitraum eine Aussage über das Verhältnis der Signale zueinander machen.<sup>44</sup>

### 4.2 Kohärenz

Zwei Signale  $a(t)$  und  $b(t)$  sind kohärent, wenn der Betrag der normierten Kreuzkorrelationsfunktion für irgendein  $\tau$  den Wert 1 annimmt.

$$|\Phi_{a,b}(\tau)| = 1.$$

Das bedeutet, die beiden Signale  $a(t)$  und  $b(t)$  sind entweder identisch, haben die gleiche Kurvenform aber eine unterschiedlich große Amplitude, sind gegenseitig verzerrungsfrei verzögert<sup>45</sup> oder verlaufen gegenphasig.

<sup>42</sup> vgl. Görne 2008, S. 131

<sup>43</sup> vgl. Blauert 1974, S. 162

<sup>44</sup> vgl. Veit 2005, S. 116

Nimmt der Korrelationskoeffizient den Wert 0 an,

$$\Phi_{x,y}(\tau) = 0.$$

so sind die beiden Signale inkohärent. In allen anderen Fällen weisen die Signale eine Teilkohärenz auf.

In der Literatur wird der Begriff der Kohärenz teilweise unterschiedlich ausgelegt. Gerne wird angenommen, dass zwei Signale nur dann kohärent sind, wenn der Korrelationskoeffizient  $\Phi_{x,y}(0) = 1$  ist - die beiden Signale also identisch und gegenseitig unverzögert sind. Dies ist aber nach der Definition der Kohärenz nicht richtig und führt oft zu Missverständnissen.

### 4.3 Addition von Amplituden und Pegeln

In der Signaladdition werden zwei Fälle der Addition unterschieden. Zum einen die Addition zweier kohärenter und gegenseitig unverzögerter Amplituden ( $\Phi_{x,y}(0) = 1$ ), zum anderen die zweier inkohärenter Amplituden ( $\Phi_{x,y}(\tau) = 0$ ). Deutlich wird dies, wenn die Addition in der Vektordarstellung veranschaulicht wird, wie in Abbildung 4.3-1 und Abbildung 4.3-2.

Generell gilt für die Vektoraddition:

$$\vec{a} + \vec{b} = \vec{c}. \quad (4.2)$$

#### 1. Addition von kohärenten und gegenseitig unverzögerten Amplituden:

Die Vektoren von Amplituden mit gleichem Phasenwinkel verlaufen parallel zueinander. Der Betrag der resultierende Amplitude  $\vec{c}$  ist dabei gleich der Summe der Beträge der beiden zu addierenden Amplituden  $\vec{a}$  und  $\vec{b}$ .<sup>46</sup>

$$|\vec{c}| = |\vec{a}| + |\vec{b}|. \quad (4.3)$$

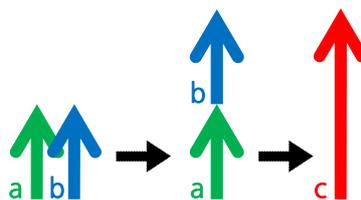


Abbildung 4.3-1 Addition zweier kohärenter und gegenseitig unverzögerter Amplituden

#### 2. Addition von inkohärenten Amplituden:

Die Vektoren inkohärenter Amplituden stehen in der Vektordarstellung orthogonal zueinander, da sie linear unabhängig sind.<sup>47</sup> Dieser Umstand führt dazu, dass die Amplitude  $\vec{c}$

<sup>45</sup> „verzerrungsfrei verzögert“ bedeutet, die beiden Signale unterscheiden sich durch eine frequenzunabhängige Phasenlaufzeit.

<sup>46</sup> vgl. Sengpiel 2003

nicht einfach über eine Addition der Beträge der Amplituden  $\vec{a}$  und  $\vec{b}$  errechnet werden kann. Der Betrag der Amplitude  $\vec{c}$  ist die Hypotenuse im rechtwinkligen Dreieck mit den Vektoren  $\vec{a}$  und  $\vec{b}$  und über den Satz des Pythagoras zu ermitteln.<sup>48</sup>

$$|\vec{c}| = \sqrt{|\vec{a}|^2 + |\vec{b}|^2}. \quad (4.4)$$

Die folgende Grafik (Abbildung 4.3-2) zeigt die Addition zweier inkohärenter Amplituden in der Vektordarstellung.

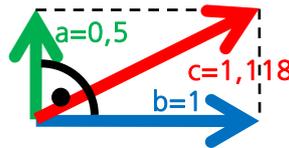


Abbildung 4.3-2 Addition inkohärenter Amplituden, wobei  $|\vec{a}| \neq |\vec{b}|$

Sollen Pegel addiert werden, so ist zu beachten, dass Pegelwerte nicht einfach aufsummiert werden dürfen. Die korrekte Vorgehensweise ist, die Pegel zuerst in ihre Amplitude umzurechnen, dann die Addition auszuführen und schließlich den entsprechenden Pegelwert wieder zu ermitteln. Dabei helfen die Formeln zur Errechnung der Amplitude A aus einem Pegel P und umgekehrt.

$$P[dB] = 20 \cdot \log_{10} \left( \frac{A}{A_0} \right), \text{ wobei } A_0 \text{ der Bezugswert ist.} \quad (4.5)$$

$$A = 10^{\frac{P[dB]}{20}} \cdot A_0. \quad (4.6)$$

Aus (4.3), (4.5) und (4.6) folgt für die Addition kohärenter und unverzögerter Signale:

$$P_c[dB] = 20 \cdot \log_{10} \left( 10^{\frac{P_a[dB]}{20}} + 10^{\frac{P_b[dB]}{20}} \right). \quad (4.7)$$

Und aus (4.4), (4.5) und (4.6) folgt für die Addition inkohärenter Signale:

$$P_c[dB] = 10 \cdot \log_{10} \left( 10^{\frac{P_a[dB]}{10}} + 10^{\frac{P_b[dB]}{10}} \right). \quad (4.8)$$

Daraus ergeben sich für die Addition von kohärenten Signalen Pegelzuwächse von bis zu 6 dB und für inkohärente Signale von bis zu 3 dB.

Aus den Definitionen folgt weiter, dass die Differenz zweier Pegel  $P_a$  und  $P_b$  das Verhältnis ihrer Amplituden repräsentiert.

<sup>47</sup> vgl. Kapitel 4.1 Korrelation

<sup>48</sup> vgl. Sengpiel 2003

$$\begin{aligned}
 P_{\text{Differenz}} &= P_a - P_b \\
 &= 20 \cdot \log_{10}(a) - 20 \cdot \log_{10}(b) \\
 &= 20 \cdot \log_{10}\left(\frac{a}{b}\right).
 \end{aligned}
 \tag{4.9}$$

Diese Gesetze gelten theoretisch sowohl für technische als auch für akustische Signale. In der Praxis ist aber bei der akustischen Addition ein Pegelzuwachs um bis zu 6 dB kaum möglich, so dass immer von einer inkohärenten Addition ausgegangen werden kann. Ein Hörer empfindet bei der Wiedergabe zweier kohärenter Signale über zwei Lautsprecher einen maximalen Pegelanstieg von 3 dB.<sup>49</sup>

#### 4.4 Kammfiltereffekt

Ein Kammfiltereffekt kann bei der Addition zweier elektrischer wie akustischer Signale entstehen. Im Folgenden werden Kammfiltereffekte betrachtet, die bei der elektrischen Signaladdition entstehen. Voraussetzung für diesen Filtereffekt ist, dass die beiden Signale kohärent oder teilkohärent und gegeneinander verzerrungsfrei verzögert, also zeitverschoben sind. Im Vergleich zu den Spektren der zu summierenden Signale entstehen abhängig von deren Zeitunterschied, Frequenz und Pegel im spektralen Verlauf des Summensignals Frequenzauslöschungen (Notches) und Pegelanhebungen (Peaks). Der Pegel der Frequenzauslöschung kann bis zu  $-\infty$  dB, die Pegelanhebung bis zu +6 dB betragen. Dieser Effekt kann eine gravierende Änderung der Klangfarbe hervorrufen. Die Stärke des Kammfilters ist dabei von der Pegeldifferenz der beiden Eingangssignale abhängig. Je kleiner die Differenz, desto stärker ist der Effekt.

Notches entstehen bei Frequenzen, deren halbe Periodendauer oder Vielfache davon der Verzögerung der beiden zu summierenden Signale entspricht.

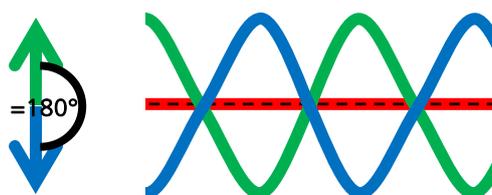


Abbildung 4.4-1 Entstehung von Notches

Die erste Notchfrequenz kann durch den Kehrwert der doppelten Zeitdifferenz  $\Delta t$  der beiden zu summierenden Signale ermittelt werden. Die weiteren Notchfrequenzen sind ungerade Vielfache der ersten Notchfrequenz, so dass gilt:

$$f_{\text{Notch}}(k)[\text{Hz}] = k \cdot \frac{1}{2 \cdot \Delta t}, \quad \text{wobei } k = 1, 3, 5, \dots$$

Peaks bilden sich zwischen den Notchfrequenzen, wo die Phasenlage der zu summierenden Signale annähernd gleich ist, also gegen  $0^\circ$  oder ein Vielfaches von  $360^\circ$  strebt.

<sup>49</sup> vgl. Veit 2005, S. 30 f.

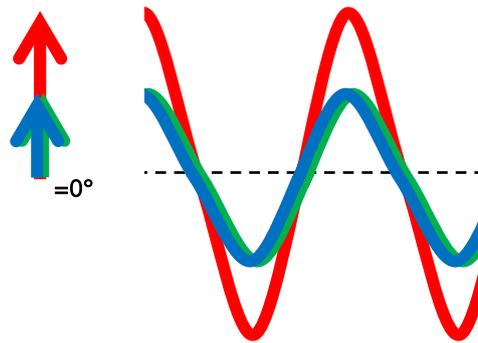


Abbildung 4.4-2 Entstehung von Peaks

Die Peakfrequenzen liegen bei geraden Vielfachen der ersten Notchfrequenz.

$$f_{Peak}(l)[Hz] = l \cdot \frac{1}{2 \cdot \Delta t}, \quad \text{wobei } l = 2, 4, 6, \dots$$

Bei einem Downmix treten Kammfilter auf, wenn zwei zu summierende Kanäle kohärente Signalanteile aufweisen, die zeitlich verschoben sind. In der Praxis passiert dies relativ häufig.

## 4.5 Phantomschallquellen

Durch die Fähigkeit des binauralen Hörens, also das Hören mit zwei Ohren, ist der Mensch in der Lage, Schallereignisse in der Horizontalebene zu orten. Ursache dafür sind hauptsächlich interaurale Laufzeit- und Pegeldifferenzen, also Differenzen der beiden Schallsignale an den Ohren. Beim natürlichen Hören stimmt in der Regel die subjektive Lokalisation der Schallquelle - das sogenannte Hörereignis - mit dem Ort der Schallerzeugung - dem Schallereignis - überein.<sup>50</sup>

Wird nun, ausgehend von der Sweet-Spot-Position, ein kohärentes Signal parallel über zwei Lautsprecher wiedergegeben, so interpretiert das Gehör bei der Lokalisation die resultierenden Ohrsignale, als würden sie von einer einzigen Schallquelle stammen. Diese Art der Wahrnehmung wird auch Summenlokalisation genannt. Die Schallquelle wird dann zwischen den beiden Lautsprechern geortet.<sup>51</sup> Die Wahrnehmung erzeugt eine fiktive Phantomschallquelle, die mehrere Schallereignisse einem Hörereignis zuordnet, so dass die Orte von Hörereignis und Schallquellen verschieden sind.<sup>52</sup> Durch Veränderung der Zeit- und/oder Pegeldifferenzen zwischen den beiden Signalen kann die Phantomschallquelle weiter zum rechten oder zum linken Lautsprecher verschoben werden.

### 4.5.1 Pegeldifferenzstereofonie

Die Pegeldifferenzstereofonie oder auch Intensitätsstereofonie ist eine von zwei Möglichkeiten, eine Phantomschallquelle zwischen zwei Lautsprechern zu verschieben. Die Auslenkung der Phantomschallquelle ist dabei von der Pegeldifferenz abhängig. In der Praxis wird dies durch den Panoramaregler am Mischpult erreicht.

<sup>50</sup> vgl. Blauert 1974, S. 163 ff.

<sup>51</sup> vgl. Görne 2008, S. 121

<sup>52</sup> vgl. Theile 1980

Helmut Wittek hat im Rahmen seiner Diplomarbeit Eigenschaften der Auslenkung von Phantomschallquellen in den von der ITU vorgeschlagenen 2.0 und 5.0 Lautsprecheraufstellungen untersucht. Umfangreiche Hörversuche ergaben, dass bei einer 2.0-Aufstellung mit einem Öffnungswinkel von  $60^\circ$  die Phantomschallquelle mit  $2,2^\circ/1$  dB Pegeldifferenz ausgelenkt wird. Bei einer 5.0 Aufstellung mit einem Öffnungswinkel von  $30^\circ$  zwischen dem Center- und dem rechten bzw. linken Lautsprecher ergeben sich  $1,1^\circ/1$  dB. Ab einer relativen Auslenkung von 75% ist das Verhältnis von Pegel und Winkel nicht mehr linear (Abbildung 4.5-1). Die volle Auslenkung wird ab einem Pegelunterschied von 18 dB erreicht, so dass die Position der Phantomschallquelle mit der des um 18 dB stärker ausgesteuerten Lautsprechers übereinstimmt.<sup>53</sup>

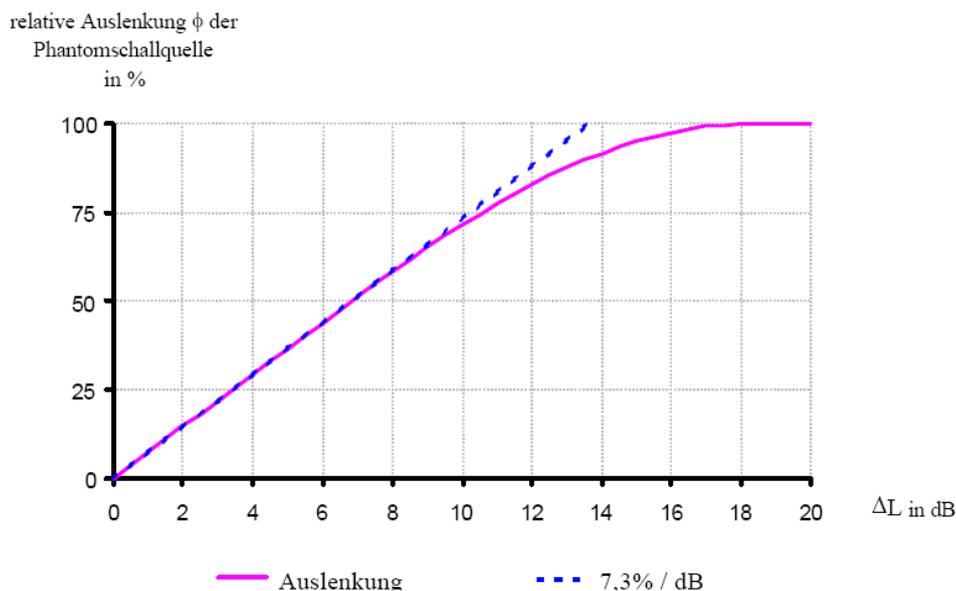


Abbildung 4.5-1 Relative Auslenkung  $\Phi$  einer Phantomschallquelle in Abhängigkeit von der Pegeldifferenz  $\Delta L$

#### 4.5.2 Laufzeitstereofonie

Die zweite Möglichkeit Phantomschallquellen zu positionieren ist die Laufzeitstereofonie. Hier werden die interauralen Differenzen durch verschiedene Laufzeiten der Signale hervorgerufen. Legt man in einer 2.0 Anordnung ein Monosignal auf beide Lautsprecher und verzögert dieses Signal in einem Kanal, so wird eine Verschiebung der Phantomschallquelle wahrgenommen.<sup>54</sup> Bei einem Öffnungswinkel von  $60^\circ$  wird eine Phantomschallquelle bis zu einer Zeitdifferenz von 0,4 ms mit  $3,8^\circ/0,1$  ms linear aus der Mitte verschoben und verändert mit zunehmender Verzögerung ihre Klangfarbe, Ausdehnung und ihren Charakter. Ab einer Laufzeitdifferenz von etwa 1 ms ist die Phantomschallquelle mit beträchtlichen Klangverfärbungen voll ausgelenkt.<sup>55</sup> Bei der Verwendung von drei Frontlautsprechern mit Öffnungswinkeln von  $30^\circ$ , wie bei der 5.0-Anordnung, bewegt sich die Phantomschallquelle mit  $1,9^\circ/0,1$  ms.

<sup>53</sup> vgl. Wittek 2000, S. 23 ff.

<sup>54</sup> vgl. Görne 2008, S. 121

<sup>55</sup> vgl. Wittek 2000, S. 23 ff.

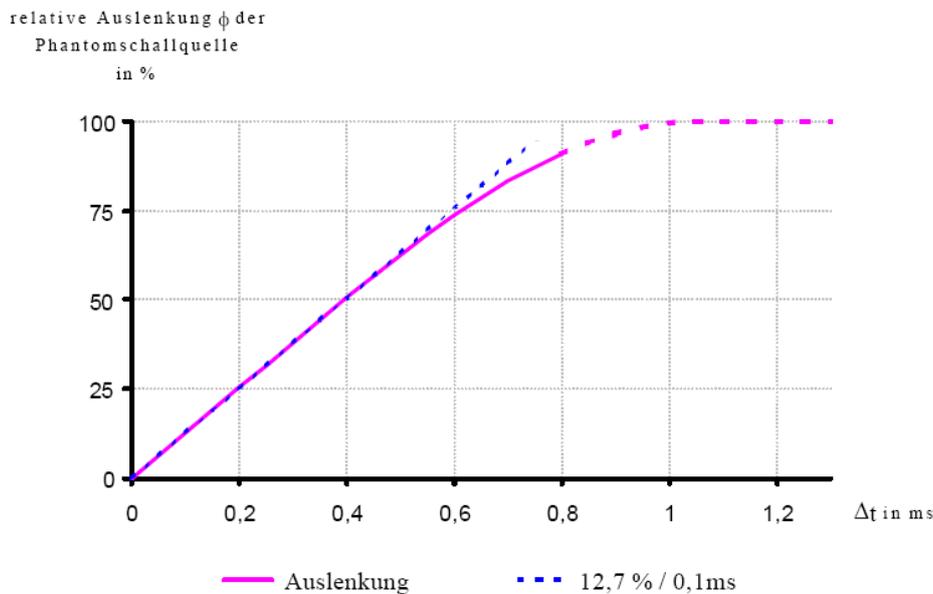


Abbildung 4.5-2 Relative Auslenkung  $\Phi$  einer Phantomschallquelle in Abhängigkeit von der Laufzeitdifferenz  $\Delta t$

Die Laufzeitstereofonie tritt in der Praxis meist nur in Kombination mit der Pegeldifferenzstereofonie auf. Mikrofonanordnungen, die eine Laufzeitdifferenz hervorrufen, wie Groß-AB<sup>56</sup> oder ORTF<sup>57</sup>, haben neben der Laufzeit- auch eine Pegeldifferenz der aufgenommenen Schallquellen zur Folge.

#### 4.6 Kopfbezogenes Koordinatensystem

Blauert führte ein kopfbezogenes Kugelkoordinatensystem ein, mit dessen Hilfe beim räumlichen Hören Ortsangaben zu einem Hörereignis oder einer Schallquelle gemacht werden können. Den dreidimensionalen Raum definierte er mit der Horizontalebene, der Frontalebene und der Medianebene. Die Lage dieser drei Ebenen kann der Abbildung 4.6-1 entnommen werden.<sup>58</sup>

<sup>56</sup> Die AB-Mikrofonanordnung sieht zwei Druck- oder Gradientenempfänger parallel nebeneinander vor. Dabei kann der Abstand variieren. Ein Abstand mit etwa 17 cm wird Klein-AB genannt - ein größerer bis zu mehreren Metern Groß-AB. Bedingt durch die Aufstellung und die Charakteristik der Mikrofone entstehen sowohl eine Pegel- als auch eine Laufzeitdifferenz. Vgl. Dickreiter, Band 1 1997, S. 286 f.

<sup>57</sup> Die ORTF-Mikrofonanordnung verwendet zwei Mikrofone mit Nieren-Charakteristik, die im Abstand von ca. 17 cm mit einem Winkel von etwa 110° zueinander aufgestellt werden. Bedingt durch die Aufstellung und die Charakteristik der Mikrofone entstehen sowohl eine Pegel- als auch eine Laufzeitdifferenz. Vgl. Dickreiter, Band 1 1997, S. 286 f.

<sup>58</sup> vgl. Blauert 1974, S. 11 f.

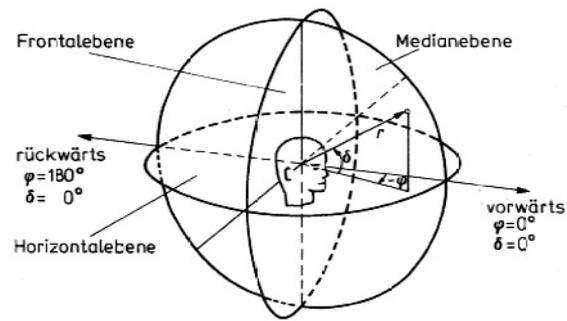


Abbildung 4.6-1 Kopfbezogenes Koordinatensystem<sup>59</sup>

## 4.7 Komplexe Zahlen

Komplexe Zahlen  $\mathbb{C}$  erweitern den Wertebereich der reellen Zahlen so, dass auch Wurzeln negativer Zahlen berechnet werden können. Dies wird durch die Einführung einer neuen Zahl  $j$  (in der Mathematik  $i$ ) realisiert, welche die Lösung der Gleichung  $x^2 = -1$  ist. Komplexe Zahlen können in zwei Formen dargestellt werden: in der algebraischen Form und in der sogenannten Polarform. Die algebraische Form trennt die komplexe Zahl  $z$  in einen reellen und einen imaginären Teil,

$$z = a + b \cdot j \quad \text{mit } a, b \in \mathbb{R}; z \in \mathbb{C}. \quad (4.10)$$

wobei der Realteil  $Re(z) = a$  und der Imaginärteil  $Im(z) = b \cdot j$  ist. In der Polardarstellung wird die komplexe Zahl  $z$  in einen Betrag  $r$  und den Phasenwinkel  $\varphi$  unterteilt.<sup>60</sup>

$$z = r \cdot e^{j\varphi}, \quad (4.11)$$

$$\text{wobei } r = \sqrt{(Re(z))^2 + (Im(z))^2}. \quad (4.12)$$

Der so konstruierte Wertebereich hat Eigenschaften, die sich vor allem in den Natur- und Ingenieurwissenschaften als äußerst dienlich erwiesen haben. Die Darstellung der komplexen Zahlenebene zeigt den Zusammenhang der algebraischen Form und der Polarform.

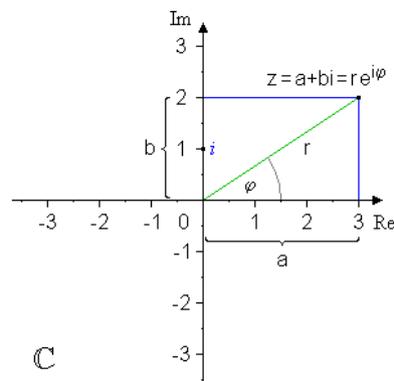


Abbildung 4.7-1 komplexe Zahlenebene

<sup>59</sup> Blauert 1974, S. 11

<sup>60</sup> Weltner, Wiesner, Heinrich, Engelhardt, Schmidt 2001, S. 183 ff.

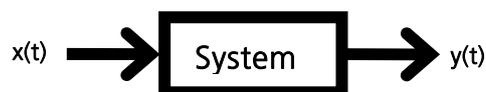
Im Gegensatz zu den reellen Zahlen decken die komplexen Zahlen einen zweidimensionalen Wertebereich ab. Die Eigenschaft, sowohl den Betrag als auch den Winkel in eine Zahl zu integrieren, prädestiniert die komplexen Zahlen zur Darstellung der DFT-Koeffizienten im Spektralbereich. Diese weisen sowohl einen Betrag - die Amplitude - als auch einen Winkel - die Phasenverschiebung - auf.

## 4.8 Digitale Signalverarbeitung

Signale als Träger von Informationen sind heute nur schwerlich wegzudenken. Gespräche mit einem Handy, Videobilder und Musik begleiten uns im Alltag. Schon 1904 definierte Meyers Konversationslexikon ein Signal als ein „bestimmtes, für das Auge (optisches Signal) oder das Ohr (akustisches Signal) berechnetes Zeichen“. Diese Begriffsbestimmung ist auch heute für elektrische und elektromagnetische Signale gültig, nur dass diese nicht mehr direkt für das Auge oder das Ohr bestimmt sind.

Elektrische Signale können als zeitkontinuierliches bzw. analoges Signal, als Abtast- bzw. zeitdiskretes Signal oder als digitales Signal auftreten.<sup>61</sup> Im Gegensatz zum zeitkontinuierlichen Signal ist das zeitdiskrete Signal nur für bestimmte Punkte definiert, in der Regel mit einem festen zeitlichen Abstand  $T$ .  $T$  ist die Abtastperiode und ihr Kehrwert die Abtastfrequenz oder Abtastrate  $f$ . Ein digitales Signal ist ein zeitdiskretes Signal, bei dem die Amplitudenwerte quantisiert wurden, also ebenfalls diskret sind.

Signale, die in elektrischer Form vorliegen, können elektronisch verarbeitet werden. Dieser Vorgang wird Signalverarbeitung genannt. Zur Veranschaulichung wird diese elektronische Verarbeitung zu einem System zusammengefasst, in dem die Signale übertragen und verändert werden (Abbildung 4.8-1). Ein sich zeitlich veränderndes Eingangssignal  $x(t)$  hat ein Ausgangssignal  $y(t)$  zur Folge, das als Systemantwort (engl.: response) bezeichnet wird.<sup>62</sup> Signalverarbeitende Systeme sind zum Beispiel Verstärker oder Filter.



**Abbildung 4.8-1** Eingangs- und Ausgangssignal eines signalverarbeitenden Systems

Unter einem linearen System versteht man ein System, welches das Eingangssignal höchstens linear verzerrt, den Pegel also frequenzabhängig oder frequenzunabhängig ändert, und die Phasenlage des Signals verschiebt. Nichtlineare Verzerrungen verändern darüber hinaus das Spektrum nicht nur quantitativ, sondern auch qualitativ, das heißt es entstehen neue Teiltöne. Ein System ist zeitinvariant, wenn seine Eigenschaften zeitunabhängig sind, das heißt wenn ein um die Zeit  $\tau$  verzögertes Eingangssignal ein ebenfalls um die Zeit  $\tau$  verzögertes Ausgangssignal hervorruft.

<sup>61</sup> vgl. Grünigen 2004, S. 1

<sup>62</sup> vgl. Görne 2008, S. 127 ff.

Ist ein System linear und zeitinvariant, so zählt es zu der wichtigsten Klasse kontinuierlicher Systeme<sup>63</sup>, den sogenannten LTI-Systemen (linearen zeitinvarianten Systemen, engl.: linear time-invariant systems).

Auf Grund ihrer Definition wird eine Schwingung  $x(t) = \cos(\omega t)$  bzw. in komplexer Schreibweise  $x(t) = e^{j\omega t}$  durch ein LTI-System höchstens in der Amplitude und der Phase verändert, so dass gilt:<sup>64</sup>

$$x(t) = e^{j\omega t} \Rightarrow y(t) = H e^{j\omega t} \quad , \text{wobei } H \in \mathbb{C}. \quad (4.13)$$

$H$  ist ein komplexer Skalierungsfaktor, der sowohl die mögliche Amplitudenänderung als auch eine mögliche Phasendrehung beinhaltet.

Besteht das Eingangssignal eines LTI-Systems aus einem Dirac-Stoß<sup>65</sup>  $x(t) = \delta(t)$ , so erhält man am Ausgang seine Impulsantwort (engl.: impulse response)  $y(t) = h(t)$ . Die Impulsantwort  $h(t)$  charakterisiert ein LTI-System vollständig und kann daher zur Systemanalyse verwendet werden, auch wenn technisch erzeugte Impulse weit vom Ideal der Dirac-Funktion entfernt sind.

Mit der Impulsantwort  $h(t)$  ist es außerdem möglich, die Ausgangsfunktion eines linearen Systems in Abhängigkeit zur Eingangsfunktion zu bestimmen. Faltet man nämlich das Eingangssignal mit der Impulsantwort, so erhält man das Ausgangssignal.<sup>66</sup>

$$y(t) = x(t) * h(t). \quad (4.14)$$

Mit dem Aufkommen leistungsfähiger, effizienter und preisgünstiger Digitalrechner, geht man mehr und mehr dazu über, Signale digital zu verarbeiten. Die Vorteile sind unter anderen die effiziente Berechnung der Fourier-Transformation und der Faltung. Liegt das Eingangssignal nicht bereits digital vor, sondern analog ( $x(t)$ ), so wird dieses nach einer analogen Vorverarbeitung (Verstärkung und Tiefpassfilterung gegen Aliasing) mit einem Analog-Digital-Wandler (AD) in ein digitales Signal überführt. Das Eingangssignal ist danach eine Folge von Zahlenwerten  $x[n]$ , die nun auf dem Digitalrechner anhand einer Rechenvorschrift zu einer Ausgangsfolge  $y[n]$  verarbeitet wird. Die Rechenvorschrift heißt Algorithmus und ist in Form eines Programms auf dem Digitalrechner implementiert. Das Verarbeiten digitaler Signale wird „digitale Signalverarbeitung“ genannt, abgekürzt „DSV“ (engl.: digital signal processing).<sup>67</sup>

Da sich diese Arbeit im Wesentlichen mit digitalen Signalen befasst, werden auch nur diese und deren digitale Verarbeitung (DSV) im Folgenden näher erläutert. Um Verwechslungen vorzubeugen, werden digitale Signale mit eckigen Klammern, analoge Signale mit runden Klammern beschrieben.

<sup>63</sup> Ein kontinuierliches System verarbeitet ein zeitkontinuierliches Eingangssignal  $x(t)$  zu einem zeitkontinuierlichen Ausgangssignal  $y(t)$ . (vgl. Grünigen 2004, S. 61)

<sup>64</sup> vgl. Görne 2008, S. 128

<sup>65</sup> Der britische Physiker und Nobelpreisträger Paul Dirac (1902-1984) entwickelte diesen „idealen Impuls“, der aus einem unendlich schmalen Impuls mit einer unendlich großen Amplitude besteht, so dass sich eine endliche Fläche mit der Größe 1 ergibt.

<sup>66</sup> vgl. Grünigen 2004, S. 62-65

<sup>67</sup> vgl. Grünigen 2004, S. 2

$$x(t) \xrightarrow{\text{AD-Wandlung}} x[n].$$

#### 4.8.1 Vom Zeit- in den Frequenzbereich

Signale werden üblicherweise als Funktionen der Zeit beschrieben. Die Information steckt damit in der zeitlichen Änderung zum Beispiel der Spannung bei einem elektrischen Signal, des Schalldrucks bei einem akustischen Signal und der zeitlichen Änderung der Lichtintensität bei einem optischen Signal.

Der Mathematikprofessor Jean Baptiste Joseph Fourier (1768-1830) entwickelte die Zerlegung eines beliebigen Signals in die im Signal vorhandenen Teilfrequenzen und deren Phasenbeziehung. Er ermöglichte mit dieser nach ihm benannten „Fourier Transformation“ die Darstellung eines Signals im Frequenzbereich (engl.: frequency domain).<sup>68</sup>

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt \quad \text{bzw.} \quad X(f) = \mathcal{F}\{x(t)\}. \quad (4.15)$$

Für die inverse Transformation gilt:

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft} df \quad \text{bzw.} \quad x(t) = \mathcal{F}^{-1}\{X(f)\}. \quad (4.16)$$

Für digitale Signale, die zeitdiskrete Folgen sind, wird das Fourier-Integral durch eine Summe approximiert. Man spricht auch von der diskreten Fourier-Transformation DFT.<sup>69</sup>

$$\mathcal{F}\{x[n]\} = X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi kn/N} \quad k = 0, 1, 2, \dots, N-1. \quad (4.17)$$

und entsprechend gilt für die inverse Transformation:

$$\mathcal{F}^{-1}\{X[k]\} = x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k]e^{j2\pi kn/N} \quad n = 0, 1, 2, \dots, N-1. \quad (4.18)$$

Das zeitdiskrete Signal  $x[n]$  ist dabei auf  $N$  Werte beschränkt. Die diskrete Fourier-Transformation liefert ein Spektrum, das ebenfalls aus  $N$  gleichmäßig über die lineare Frequenzskala verteilten Werten besteht, den sogenannten DFT-Koeffizienten (auch DFT-bins genannt). Daher steigt auch die Genauigkeit des Spektrums mit der Anzahl  $N$  der Koeffizienten.<sup>70</sup> Der Abstand  $\Delta f$  zweier DFT-Koeffizienten wird Frequenzauflösung genannt.

Mittels der DFT ist es möglich, für jedes beliebige Signal das dazugehörige Spektrum zu errechnen. Da diese Rechnung jedoch nicht für unendlich viele, sondern nur für  $N$  Werte durchgeführt wird, ist die DFT generell fehlerbehaftet. Möchte man ein zeitlich unbegrenztes oder relativ langes Signal in den Spektralbereich überführen, muss dieses vor der Transformation in geeignet große

<sup>68</sup> vgl. Grünigen 2004, S. 28

<sup>69</sup> vgl. Kammeyer & Kroschel 2006, S. 16 ff.

<sup>70</sup> vgl. Werner 2006, S. 50 ff.

Fenster unterteilt werden.<sup>71</sup> Die Dauer des Fensters  $NT$  heißt Messdauer, Messintervall oder Fensterlänge. Der Kehrwert der Messdauer  $NT$  ergibt die Frequenzauflösung  $\Delta f$ .<sup>72</sup>

$$\Delta f = \frac{1}{NT} \quad (4.19)$$

Wegen der Symmetrie des gespiegelten diskreten Spektrums<sup>73</sup> beschreibt schon die Hälfte der berechneten Werte das Spektrum des Signals eindeutig. Eine DFT mit der Länge von 2048 Samples erreicht effektiv eine Aufgliederung des Spektralbereichs in nur 1024 Samples.<sup>74</sup>

Um die DFT in Echtzeit zu realisieren, schlugen James Cooley (\*1926) und John W. Tukey (1915-2000) im Jahr 1965 ein Verfahren vor, das die Berechnung der DFT stark beschleunigte. Die schnelle Fourier-Transformation (FFT, engl.: fast fourier transformation) kann die DFT durch die geschickte Anwendung verschiedener Verfahren, wie zum Beispiel die Zweierkomplement-Codierung, deren Ansätze bis auf Gauß (1805) zurückreichen, um mehr als hundertmal schneller berechnen. Allerdings erwartet der FFT-Algorithmus ein Eingangssignal mit der Länge einer Zweierpotenz ( $N = 2^n$ ).<sup>75</sup> Daher werden typischerweise Längen von  $N = 512, 1024, 2048, 4096$  oder  $8192$  Samples verwendet.

Mit Hilfe der Fourier-Transformation (4.15) kann theoretisch jede Funktion aus dem Zeitbereich, als eine Funktion im Spektralbereich dargestellt werden. Aus dieser Überlegung folgt für ein LTI-System:

$$x(t) \stackrel{DFT}{\iff} X(f),$$

$$h(t) \stackrel{DFT}{\iff} H(f),$$

$$y(t) \stackrel{DFT}{\iff} Y(f).$$

Die Faltung im Zeitbereich entspricht einer Multiplikation im Spektralbereich.<sup>76</sup> Somit gilt:

$$y(t) = x(t) * h(t) \stackrel{DFT}{\iff} Y(f) = X(f) \cdot H(f). \quad (4.20)$$

Diese beiden Gleichungen beschreiben die fundamentale Beziehung zwischen dem Eingangssignal und dem Ausgangssignal eines LTI-Systems. Dabei ist  $H(f)$  die Übertragungsfunktion des Systems, auch Frequenzgang (engl.: frequency response) genannt.

<sup>71</sup> vgl. Werner 2006, S. 53 f.

<sup>72</sup> vgl. Grünigen 2004, S. 169

<sup>73</sup> Bei der Abtastung im Zuge der AD-Wandlung entsteht eine Impulsamplitudenmodulation. Die Amplitudenmodulation besitzt die Eigenschaft, dass im Spektralbereich neben dem Träger zwei Seitenbänder entstehen. Damit diese Seitenbänder sich nicht mit dem Träger überschneiden (Aliasing), muss das Signal mit einer Frequenz abgetastet werden, die mindestens doppelt so groß ist wie die höchste zu übertragende Frequenz (Shannon-Abtasttheorem). Die Folge ist ein gespiegeltes Spektrum.

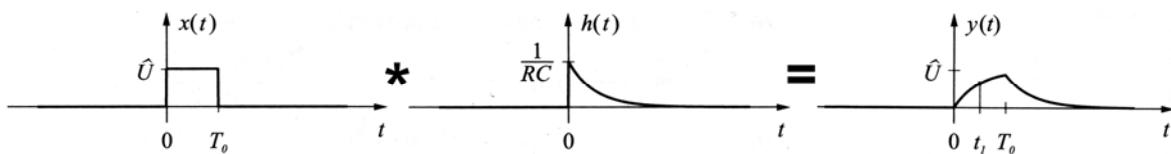
<sup>74</sup> vgl. Grünigen 2004, S. 170

<sup>75</sup> vgl. Wendemuth 2005, S. 165 ff.

<sup>76</sup> vgl. Kammeyer & Kroschel 2006, S. 19 f.

Das Ausgangssignal eines LTI-Systems kann damit sowohl im Zeitbereich, durch Faltung mit der Impulsantwort  $h(t)$ , als auch im Spektralbereich, durch Multiplikation mit der Übertragungsfunktion  $H(f)$ , berechnet werden. Die „schnelle Faltung“ (engl.: fast convolution) bedient sich dieser Möglichkeit unter Verwendung der FFT und kann dadurch den Rechenaufwand einer Faltung deutlich verringern.<sup>77</sup>

Auch bei dem IRT-Downmix wird dieser Trick angewandt und die DFT-Koeffizienten im Spektralbereich mit Variablen multipliziert. Dabei muss berücksichtigt werden, dass durch die im Zeitbereich resultierende Faltung das Ausgangssignal im Vergleich zu den Eingangssignalen länger wird<sup>78</sup> - sozusagen eine Art Ein- bzw. Ausschwingvorgang erhält (Abbildung 4.8-2).



**Abbildung 4.8-2** Grafische Darstellung einer Faltung  $x(t) * h(t) = y(t)$

Mit der digitalen Signalverarbeitung kann für den Downmix eine aktive Bearbeitung erreicht werden, also ein intelligentes Verfahren, das auf die Eingangssignale mit dynamischen Parametern reagieren kann.

<sup>77</sup> vgl. Kammeyer & Kroschel 2006, S. 248 ff.

<sup>78</sup> Das Faltungsprodukt ist so lang wie die Summe der Längen der beiden zu faltenden Signale minus eins. (vgl. Werner 2006, S. 75 ff.) Um die so entstehende zusätzliche Information nicht zu verlieren, kommt die Anwendung des Zeropaddings (vgl. Kapitel 6.2.1 Ablauf des Downmixvorgangs; 7.2.5 Ein- und Ausschwingvorgang durch die Bearbeitung) gelegen. Die zusätzliche Information wird in den mit Nullen belegten Teil des Fensters geschrieben und wird so beim Zusammensetzen der Fenster korrekt in das resultierende Ausgangssignal integriert.

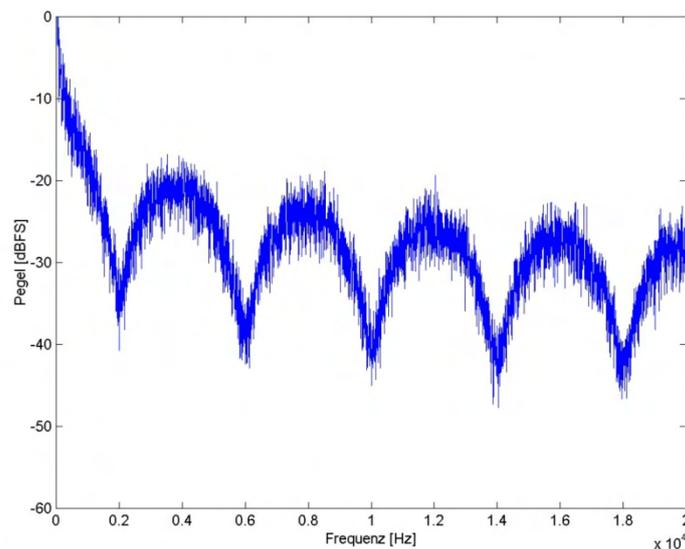
## 5 Probleme und Fehler, die beim Stereodownmix auftreten

Wie bereits in den vorhergehenden Kapiteln angedeutet, kann der Downmixvorgang in einigen Fällen zu Problemen und sogar zu Fehlern führen. Ein neuer Algorithmus soll auf diese Fälle eingehen können, sie kompensieren oder verhindern. Bevor nach Lösungen gesucht werden kann, sind die Probleme und Fehler im Detail zu betrachten.

### 5.1 Laufzeitunterschiede

Bei der Matrixierung kohärenter oder teilkohärenter Signale mit Phasenverschiebungen - also unterschiedlichen Laufzeiten - kommt es zu Interferenzen, welche Pegelanhebungen, Pegelabsenkungen oder sogar vollständige Auslöschungen zur Folge haben, die Charakteristika eines Kammfilters.<sup>79</sup>

Die Ursache für Laufzeitunterschiede liegt meist aufnahmeseitig in der Methode der Mikrofonierung und Klanggestaltung. Beispiele dafür sind gezielte stereophone Aufnahmen mittels Laufzeitstereofonie, die Verwendung mehrerer Stützmikrofone, die Verwendung zeitverzögernder Effekte, Laufzeit-Panning einer Phantomschallquelle<sup>80</sup> oder der im Mehrkanalton beliebte und bewusst angewendete Präzedenzeffekt<sup>81</sup>. Verbindliche Empfehlungen, sich nur hinsichtlich des Downmixes unkritischer Methoden zu bedienen, wie zum Beispiel Mikrofonanordnungen, die ein möglichst geringes Übersprechen bewirken und somit nahezu inkohärente Signale ergeben, machen erfahrungsgemäß wenig Sinn. Sie würden Tonmeister und Toningenieur in ihrem kreativen Gestaltungsspielraum unvermeidbar einschränken. Somit bleibt nur die Möglichkeit, die auftretenden Kammfilter beim Downmixvorgang zu kompensieren.



**Abbildung 5.1-1** Kammfilter nach einem ITU-Downmix bei  $\Delta t_{R/RS} = 0,25 \text{ ms}$

<sup>79</sup> vgl. 4.4 Kammfiltereffekt; vgl. 4.5.2 Laufzeitstereofonie

<sup>80</sup> vgl. 4.5 Phantomschallquellen

<sup>81</sup> auch „Gesetz der ersten Wellenfront“ genannt, welches besagt, dass der Hörer, wenn mehrere Schallsignale unterschiedlich zeitverzögert (mindestens 1 ms) aus verschiedenen Richtungen bei ihm eintreffen, das Schallsignal aus der Richtung des zuerst eingetroffenen Signals wahrnimmt. (vgl. Blauert 1974, S. 163)

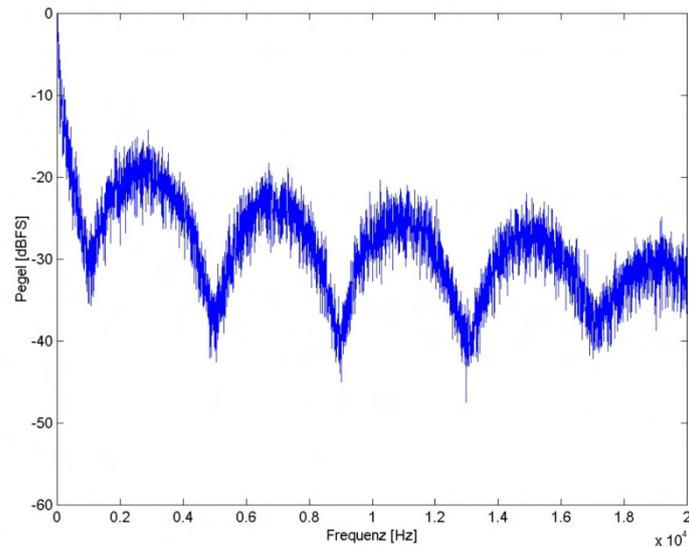


Abbildung 5.1-2 Kammfilter nach einem Dolby Pro Logic II-Downmix bei  $\Delta t_{R/RS} = 0,25 \text{ ms}$

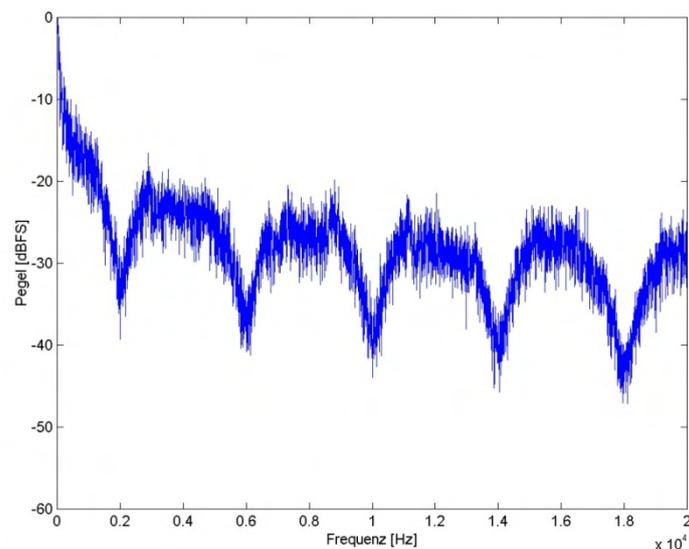


Abbildung 5.1-3 Kammfilter nach einem C.T.-Downmix bei  $\Delta t_{R/RS} = 0,25 \text{ ms}$

Betrachtet man die Kammfilter, die künstlich beim Downmix eines Mehrkanalsignals hervorgerufen werden, das auf R und RS rosa Rauschen<sup>82</sup> mit einer Zeitdifferenz von 25 ms enthält, so erkennt man unschwer in Abbildung 5.1-1 und Abbildung 5.1-2 die ausgeprägten Peaks und Notches der Downmixes mit der ITU-Empfehlung und Dolby Pro Logic II. Das Frequenzspektrum des C.T.-Downmixes weist hingegen eine schwache Kompensation des Kammfilters auf (Abbildung 5.1-3).

<sup>82</sup> Rosa Rauschen bezeichnet ein Rauschen, deren Rauschleistungsdichte mit dem Anstieg der Frequenz abnimmt.

## 5.2 Veränderung der effektiven Energie

Bei einem Downmixvorgang ist es notwendig, den Center- und die Surroundkanäle, die bei einer zweikanaligen Abbildung nicht mehr vorhanden sind, zu einem sinnvollen Anteil in den linken und rechten Kanal der Stereoverversion zu integrieren. Hierbei ist wichtig, die Abbildung, die Lautstärke sowie die Lautstärkebalance zu erhalten.

Fast alle Downmixverfahren, allen voran die ITU-Empfehlung, verringern vor der Matrixierung den Pegel des Centerkanals um 3 dB (Abbildung 3.2-1, Abbildung 3.2-2 und Abbildung 3.3-1). In den resultierenden Seitenkanälen entstehen kohärente Signalanteile, die bei der Wiedergabe eine mittige Phantomschallquelle als Ersatz für den Centerlautsprecher erzeugen. Durch die akustische Addition der beiden Seitenkanäle erlangt diese Phantomschallquelle auch theoretisch wieder die akustische Leistung des Centerkanals. Jedoch ist diese Aufteilung des Centerkanals nicht in allen Fällen zufriedenstellend. Drei theoretische Beispiele sollen die Problematik verdeutlichen.

In diesem und den folgenden Beispielen werden die elektrischen Signalpegel in dBFS (Dezibel Full Scale; unter Bezug auf die digitale Vollaussteuerung) und die Änderung des Schalldruckpegels (dB SPL, Dezibel Sound Pressure Level) zwischen der Mehrkanal- und der Stereoverversion bei der Wiedergabe über Lautsprecher in dB angegeben.

### 5.2.1 Erstes Beispiel

Für das erste Beispiel sei angenommen, dass L und C kohärente Signale ohne Phasenverschiebung und mit einem Pegel von -6 dBFS enthalten.

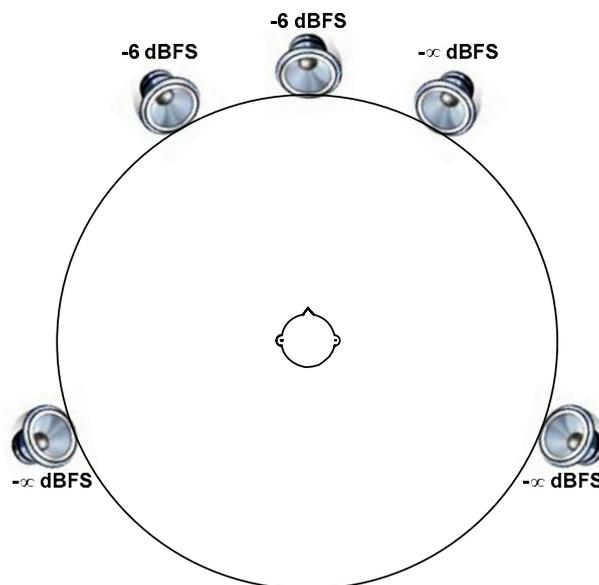


Abbildung 5.2-1 5.0 Abhörsituation mit  $L = C = -6$  dBFS.

Nach dem ITU-Downmix ergeben sich unter Verwendung der Gleichung (4.7) für die Addition elektrischer Signale folgende Pegel:

$$L_{ITU} = 20 \cdot \log_{10} \left( 10^{\frac{-6 \text{ dBFS}}{20}} + 10^{\frac{-6 \text{ dBFS} - 3 \text{ dBFS}}{20}} \right) = -1,35 \text{ dBFS},$$

$$R_{ITU} = -6 \text{ dBFS} - 3 \text{ dBFS} = -9 \text{ dBFS}.$$

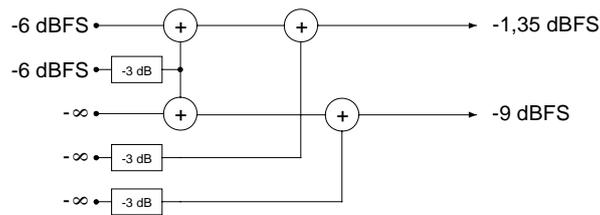


Abbildung 5.2-2 Ein- und Ausgangspegel nach einem ITU-Downmix mit  $C = L = -6$  dBFS

Die Änderung des Schalldruckpegels am Sweet-Spot kann durch die Differenz der für die Mehrkanal- und die Stereoverversion mit (4.8) bestimmten Summenpegel errechnet werden:

$$P_{\text{Differenz}} = |(-3 \text{ dBFS})_{5,0} - (-0,66 \text{ dBFS})_{2,0}| = 2,34 \text{ dB}.$$

Ein solcher Pegelunterschied ist hörbar.

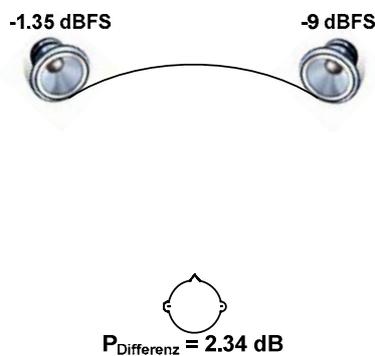


Abbildung 5.2-3 2.0 Abhörsituation nach einem ITU-Downmix mit  $L = C = -6$  dBFS

## 5.2.2 Zweites Beispiel

Für das zweite Beispiel wird angenommen, dass das gleiche kohärente Signal ohne Phasenverschiebung auf den Kanälen L und LS mit einem Pegel von -6 dBFS anliegt (Abbildung 5.2-4).

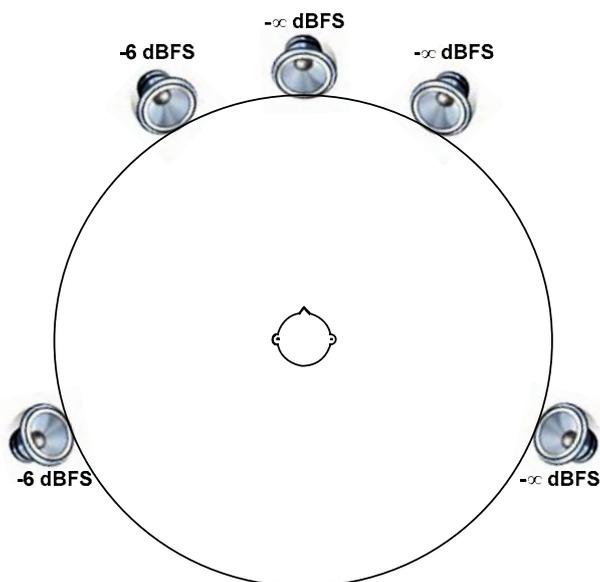
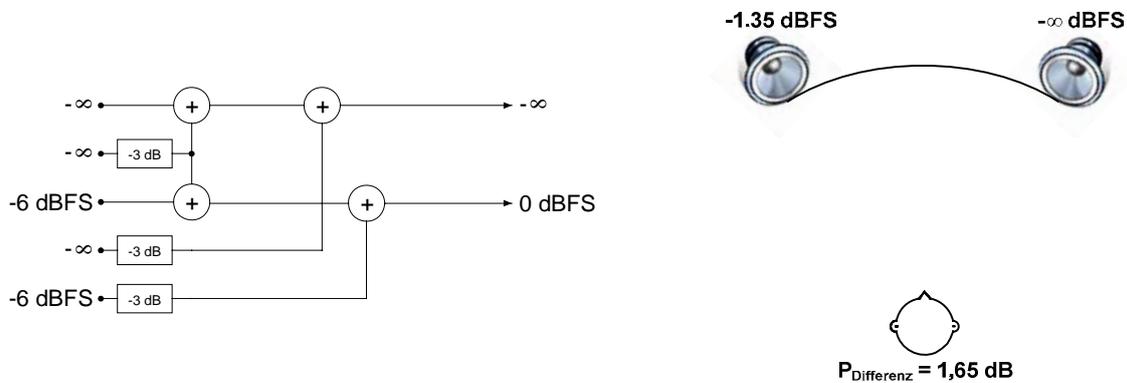


Abbildung 5.2-4 5.0 Abhörsituation mit  $L = LS = -6$  dBFS

Nach der ITU-Empfehlung ergeben sich dafür folgende Ausgangspegel:

$$L_{ITU} = 20 \cdot \log_{10} \left( 10^{\frac{-6 \text{ dBFS}}{20}} + 10^{\frac{-6 \text{ dBFS} - 3 \text{ dBFS}}{20}} \right) = -1,35 \text{ dBFS},$$

$$R_{ITU} = -\infty.$$



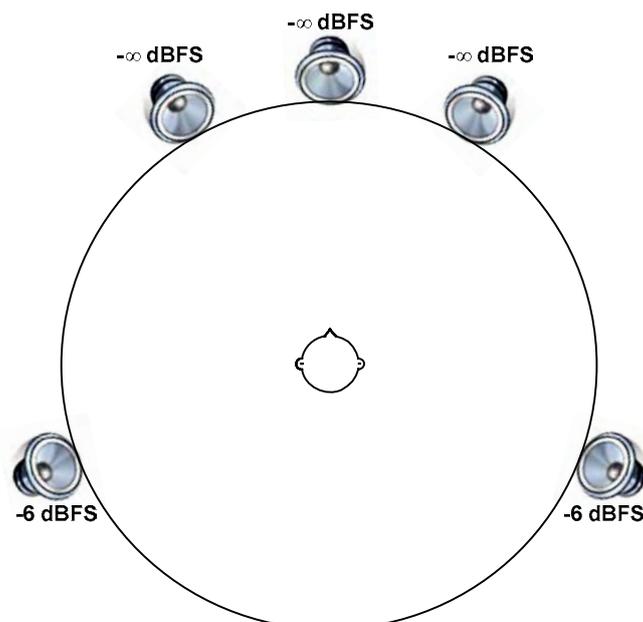
**Abbildung 5.2-5** Ein- und Ausgangspegel nach einem ITU-Downmix mit  $L = LS = -6 \text{ dBFS}$

**Abbildung 5.2-6** 2.0 Abhörsituation nach einem ITU-Downmix mit  $L = LS = -6 \text{ dBFS}$

An der Abhörposition resultiert eine Änderung des Schalldruckpegels bei der Wiedergabe der Downmixversion im Vergleich zu dem des Mehrkanaltons von  $P_{Differenz} = 1,65 \text{ dB}$ .

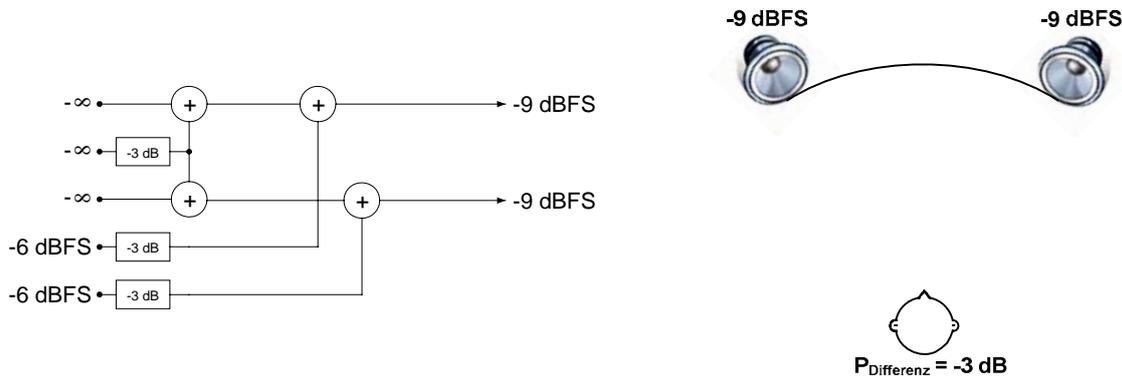
### 5.2.3 Drittes Beispiel

Die beiden pegelgleichen, kohärenten und nicht phasenverschobenen Signale liegen jetzt an LS und RS an, so dass sich die in Abbildung 5.2-7 dargestellte Situation ergibt.



**Abbildung 5.2-7** 5.0 Abhörsituation mit  $LS = RS = -6 \text{ dBFS}$

Nach dem Downmix gemäß der ITU-Empfehlung liegen die beiden Signale mit einer Dämpfung von 3 dB auf L und R.



**Abbildung 5.2-8** Ein- und Ausgangspegel nach einem ITU-Downmix mit  $LS = RS = -6$  dBFS

**Abbildung 5.2-9** 2.0 Abhörposition nach einem ITU-Downmix mit  $LS = RS = -6$  dBFS

Die Phantomschallquelle, die durch das LS- und RS-Signal mittig im Rücken des Hörers entsteht, wandert nach dem Downmix in die Mitte nach vorne. Der Schalldruckpegel ist nach dem Downmix an der Abhörposition um 3 dB niedriger. Die Dämpfung der rückwärtigen Signale beim Downmix hat jedoch im Vergleich zur Dämpfung der vorderen Kanäle oft einen positiven Effekt. Bei klassischer Musik enthalten die beiden Surroundkanäle im Wesentlichen die akustische Abbildung des Raumes. Wird dieser Hall mit voller Leistung auf den linken und rechten Kanal des Downmixes gemischt, so ist das Klangbild sehr schnell „schwammig“ und verliert an Präsenz. Bei einem Handmix der Stereoverision würde der Hall sparsamer eingesetzt, so dass die Dämpfung der rückwärtigen Kanäle um 3 dB ein sinnvolles Ergebnis liefern dürfte.

Auch bei Popmusik kommt diese Absenkung nicht ungelegen. Zwar lässt sich hier schwerer eine generelle Aussage treffen, dennoch verwenden viele Produktionen die Kanäle LS und RS für zusätzliche Effekte und Sounds, die beim Zumischen auf die Downmixkanäle das Klangbild überladen und wichtige Signalanteile verdecken können.<sup>83</sup> Eine Dämpfung ist in einem solchen Fall von Vorteil.

Lediglich Produktionen, in welchen die rückwärtigen Kanäle auch Direktschall enthalten, wie zum Beispiel Hörspiele, in denen eine Person von hinten spricht, sind durch den resultierenden Schallpegelunterschied benachteiligt. An dieser Stelle wäre in einem Downmix eine Fallunterscheidung abhängig vom Genre sinnvoll, ähnlich wie sie der ORF vornimmt.<sup>84</sup> Die Dämpfung der Surroundkanäle kann dann abhängig vom Ausgangsmaterial eingestellt werden.

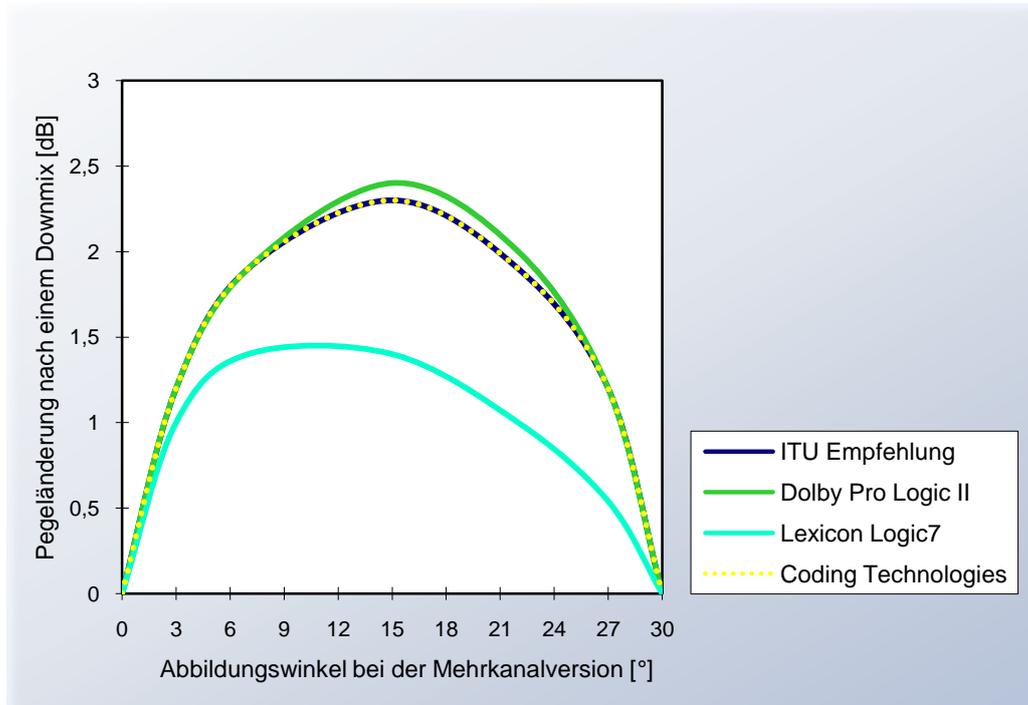
#### 5.2.4 Ergebnis

Auch wenn das dritte Beispiel in der Praxis eher unproblematisch sein dürfte, so haben die ersten beiden Beispiele gezeigt, dass Phantomschallquellen, die mit Hilfe des Centerkanals oder eines Surroundkanals im vorderen Halbkreis erzeugt werden, ihre effektive Energie während des Downmixvorgangs nicht halten können.

<sup>83</sup> vgl. Kapitel 10.1 Lautheit und Dynamik von Stereo- und Mehrkanal-Produktionen

<sup>84</sup> vgl. Kapitel 3.2.1 ITU Downmix

Mit dieser Problematik haben auch die anderen Downmixverfahren zu kämpfen. Abbildung 5.2-10 zeigt die resultierende Pegeländerung einer Phantomschallquelle nach verschiedenen Downmixverfahren in Bezug auf ihre Position in der Mehrkanalversion. Das Verfahren von Lexicon schneidet dabei auf Grund seiner Energieanpassung am besten ab.



**Abbildung 5.2-10** Pegelzunahme nach einem Downmix in Abhängigkeit vom Abbildungswinkel einer Phantomschallquelle in der Mehrkanalversion.

### 5.3 Differenzpegeländerung

Ein Downmix verändert nicht nur den Gesamtpegel im Vergleich zur Mehrkanalversion, sondern auch die Pegeldifferenz der kohärenten und inkohärenten Signalanteile. Dies führt bei der resultierenden Stereoverversion zu einer nicht unerheblichen Änderung in der Lautstärkebalance.

Die Differenz der Pegel der kohärenten und inkohärenten Signalanteile wird in dieser Arbeit Differenzpegel genannt, die Änderung des Differenzpegels von der Mehrkanalversion im Vergleich zur herunter gemischten Zweikanalversion wird Differenzpegeländerung genannt.

Das folgende Beispiel soll die Differenzpegeländerung anschaulich machen. Die kohärenten und inkohärenten Signalanteile zweier Kanäle des Mehrkanaltons addieren sich bei der Wiedergabe akustisch derart, dass sich sowohl die kohärenten als auch die inkohärenten Signalanteile um bis zu 3 dB anheben. Im Beispiel liegt sowohl im Center- als auch im linken Signal der Pegel des inkohärenten Signalanteils bei -6 dBFS, der Pegel des kohärenten Signalanteils bei -4 dBFS. Am Sweet-Spot resultiert daraus ein Differenzpegel von:

$$P_{DP\ 5.0} = |(-3\ dBFS)_{inkohärent} - (-1\ dBFS)_{kohärent}| = 2\ dB.$$

Abbildung 5.3-1 veranschaulicht diese Ausgangssituation.

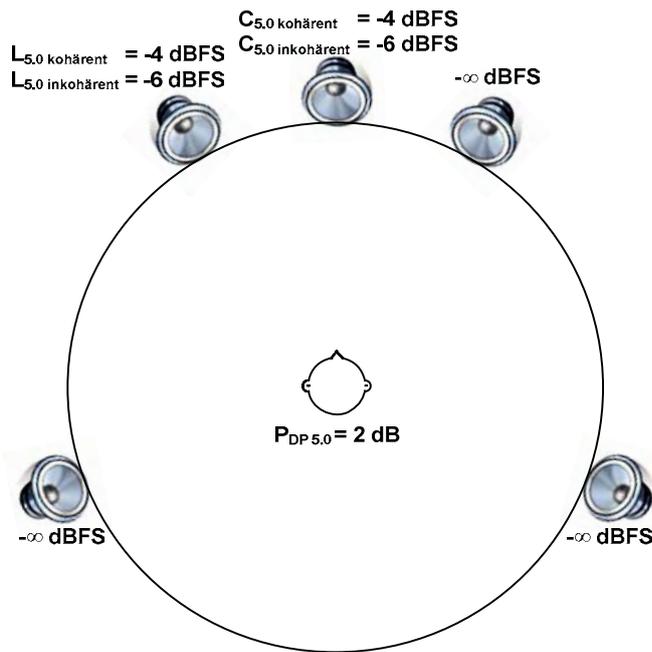


Abbildung 5.3-1 5.0 Abhörsituation mit einem Differenzpegel von  $P_{DP\ 5.0} = 2\ dB$ .

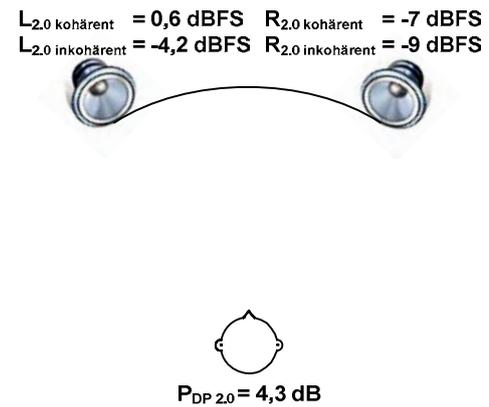


Abbildung 5.3-2 2.0 Abhörsituation mit einem Differenzpegel von  $P_{DP\ 2.0} = 4,3\ dB$ .

Beim Downmix wird das linke Signal mit dem um 3 dB abgesenkten Center-Signal elektrisch addiert. Aus (4.7) und (4.8) ergeben sich für den inkohärenten bzw. kohärenten Signalanteil die folgenden Pegel:

$$L_{\text{inkohärent}} = 10 \cdot \log_{10} \left( 10^{\frac{-6\ \text{dBFS}}{10}} + 10^{\frac{-6\ \text{dBFS} - 3\ \text{dBFS}}{10}} \right) = -4,2\ \text{dBFS},$$

$$L_{\text{kohärent}} = 20 \cdot \log_{10} \left( 10^{\frac{-4\ \text{dBFS}}{20}} + 10^{\frac{-4\ \text{dBFS} - 3\ \text{dBFS}}{20}} \right) = 0,6\ \text{dBFS}.$$

Auf dem rechten Downmixkanal liegt nur das gedämpfte Center-Signal an.

$$R_{\text{inkohärent}} = -6\ \text{dBFS} - 3\ \text{dBFS} = -9\ \text{dBFS},$$

$$R_{\text{kohärent}} = -4\ \text{dBFS} - 3\ \text{dBFS} = -7\ \text{dBFS}.$$

Daraus resultiert ein Differenzpegel am Sweet-Spot zwischen den kohärenten und inkohärenten Signalanteilen von  $P_{DP\ 2.0} = 4,3\ \text{dB}$  (Abbildung 5.3-2). Es entsteht im Vergleich zum Differenzpegel der 5.0-Version eine Differenzpegeländerung von:

$$P_{\text{Änderung}} = |(2\ \text{dB})_{5.0} - (4,3\ \text{dB})_{2.0}| = 2,3\ \text{dB},$$

die gut hörbar ist.

Der Downmix verändert die Pegeldifferenz kohärenter und inkohärenter Signalanteile, führt also zu einer Differenzpegeländerung und verfälscht somit die ursprüngliche Lautstärkebalance.

## 5.4 Phantomschallquellenverschiebung

Als letzter Punkt ist die Phantomschallquellenverschiebung zu nennen. Nicht nur die effektive Energie einer Phantomschallquelle verändert sich beim Downmix,<sup>85</sup> sondern auch deren Auslenkung aus der Mitte, wie die folgenden Beispiele zeigen.

### 5.4.1 Phantomschallquellen durch Pegeldifferenzen

In Anlehnung an das erste Beispiel in Kapitel 5.2.1 wird die Änderung des Auslenkungswinkels  $\alpha$  während eines Downmixvorgangs untersucht. Die kohärenten, nicht phasenverschobenen und pegelidentischen Signale auf den Kanälen L und C bilden mittig zwischen den beiden Lautsprechern eine Phantomschallquelle, die mit einem Winkel  $\alpha_{5,0} = 15^\circ$  in der Horizontalebene vom Center- in Richtung des linken Lautsprechers ausgelenkt ist (Abbildung 5.4-1).

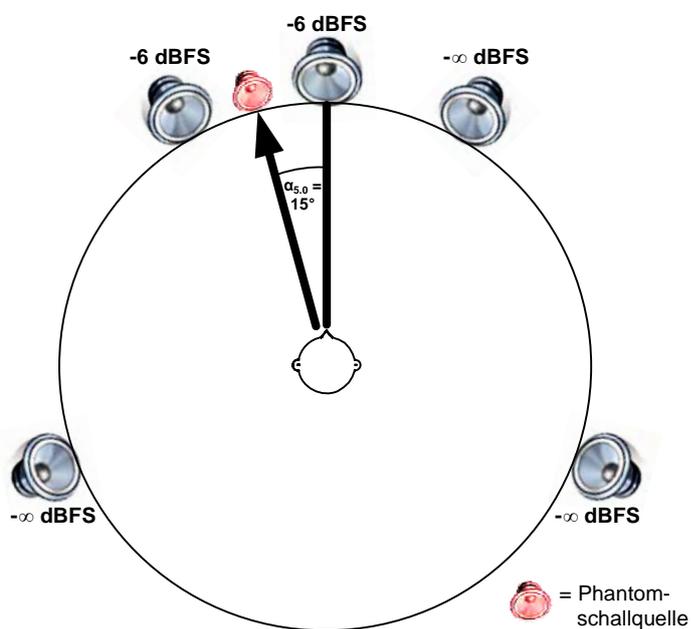


Abbildung 5.4-1 5.0 Abhörsituation mit Phantomschallquelle bei L = -6 dBFS und C = -6 dBFS.

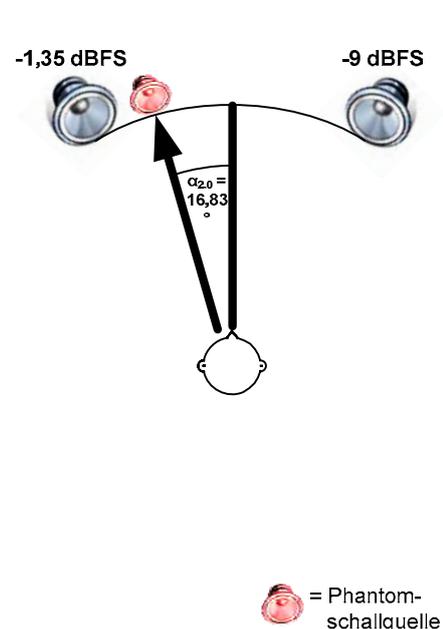


Abbildung 5.4-2 2.0 Abhörsituation nach einem ITU-Downmix mit L = -1,35 dBFS und R = -9 dBFS.

Analog zu Kapitel 5.2.1 erhält man beim Downmix die Ausgangspegel  $L_{ITU} = -1,35 \text{ dBFS}$  und  $R_{ITU} = -9 \text{ dBFS}$ . Die Pegeldifferenz der beiden Stereokanäle gibt Auskunft über die Auslenkung nach dem Downmix.

$$\Delta LR = -9 \text{ dBFS} - (-1,35 \text{ dBFS}) = 7,65 \text{ dBFS}$$

Wie in Kapitel 4.5.1 gezeigt wurde, wird eine Phantomschallquelle in einer Stereoanordnung mit etwa  $2,2^\circ/1 \text{ dB}$  ausgelenkt.

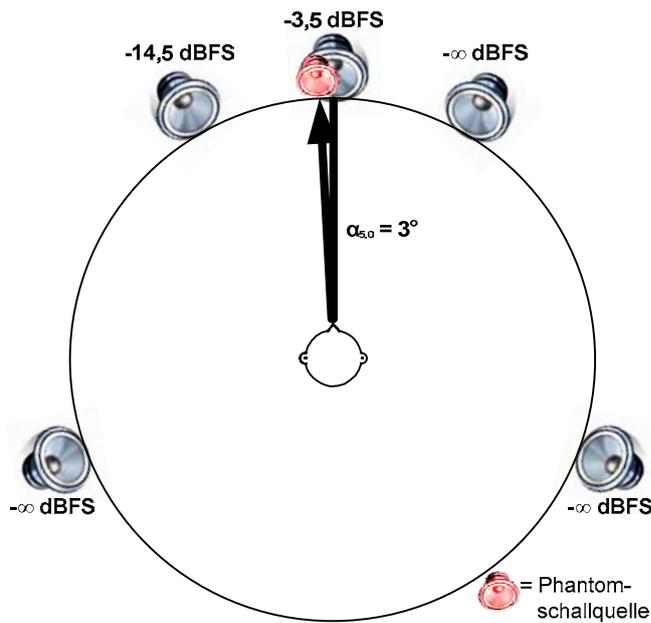
$$\alpha_{2,0} = 7,65 \text{ dB} \cdot 2,2^\circ/\text{dB} = 16,83^\circ$$

$$\Delta\alpha = \alpha_{5,0} - \alpha_{2,0} = 15^\circ - 16,83^\circ = 1,83^\circ$$

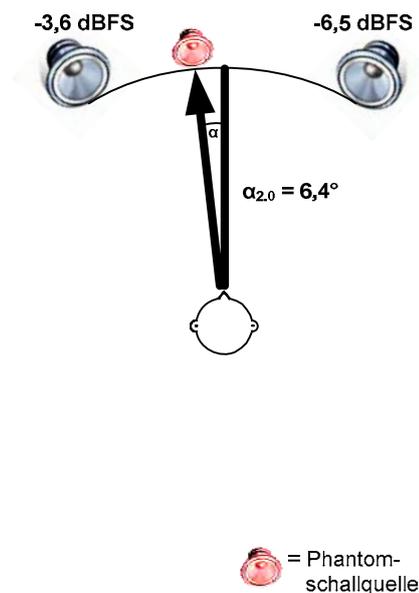
<sup>85</sup> vgl. 5.2 Veränderung der effektiven Energie

Im Vergleich zur Wiedergabe des Mehrkanaltons entstehen eine Winkeldifferenz und damit eine Verschiebung der Phantomschallquelle um  $1,83^\circ$  in der Horizontalebene aus der Mitte heraus. In Abbildung 5.4-1 und Abbildung 5.4-2 ist diese Verschiebung dargestellt.

In einem zweiten Beispiel sei angenommen, dass L und C kohärente und phasengleiche Signale enthalten, der linke Kanal jedoch mit  $-14,5$  dBFS und der Centerkanal mit  $-3,5$  dBFS angesteuert ist. Aus einer Pegeldifferenz von  $11$  dB zwischen L und C resultiert eine Phantomschallquelle, die vom Centerlautsprecher um  $\alpha_{5,1} = 3^\circ$  in Richtung des linken Lautsprechers ausgelenkt ist (Abbildung 5.4-3).



**Abbildung 5.4-3** 5.0 Abhörsituation mit Phantomschallquelle bei  $L = -14,5$  dBFS und  $C = -3,5$  dBFS.



**Abbildung 5.4-4** 2.0 Abhörsituation nach einem ITU-Downmix mit  $L = -3,6$  dBFS und  $R = -6,5$  dBFS.

Nach dem ITU-Downmix erhält man unter Verwendung der Gleichung (4.7) für die Addition elektrischer, kohärenter und unverzögerter Signale folgende Pegel:

$$L_{ITU} = 20 \cdot \log_{10} \left( 10^{\frac{-14,5 \text{ dBFS}}{20}} + 10^{\frac{-3,5 \text{ dBFS} - 3 \text{ dBFS}}{20}} \right) = -3,6 \text{ dBFS},$$

$$R_{ITU} = -3,5 \text{ dBFS} - 3 \text{ dBFS} = -6,5 \text{ dBFS}.$$

Zwischen  $L_{ITU}$  und  $R_{ITU}$  ergibt dies eine Pegeldifferenz von  $\Delta LR = 2,9 \text{ dB}$ , die eine Auslenkung von  $\alpha_{2,0} = 2,9 \text{ dB} \cdot 2,2^\circ/\text{dB} = 6,4^\circ$  in der Horizontalebene in Richtung des linken Lautsprechers bewirkt (Abbildung 5.4-4). Es entsteht ein Darstellungsfehler von  $\Delta\alpha = 6,4^\circ - 3^\circ = 3,4^\circ$ .

Die räumliche Auflösung des menschlichen Gehörs zeigt bei der Ortung von Hörereignissen Schwächen. Blauert definiert diese als Lokalisationsunschärfe. Abhängig von Pegeldifferenz und Frequenz bzw. Frequenzspektrum ist eine Phantomschallquelle unterschiedlich genau ortbar.<sup>86</sup>

<sup>86</sup> vgl. Blauert 1974, S. 30 ff.

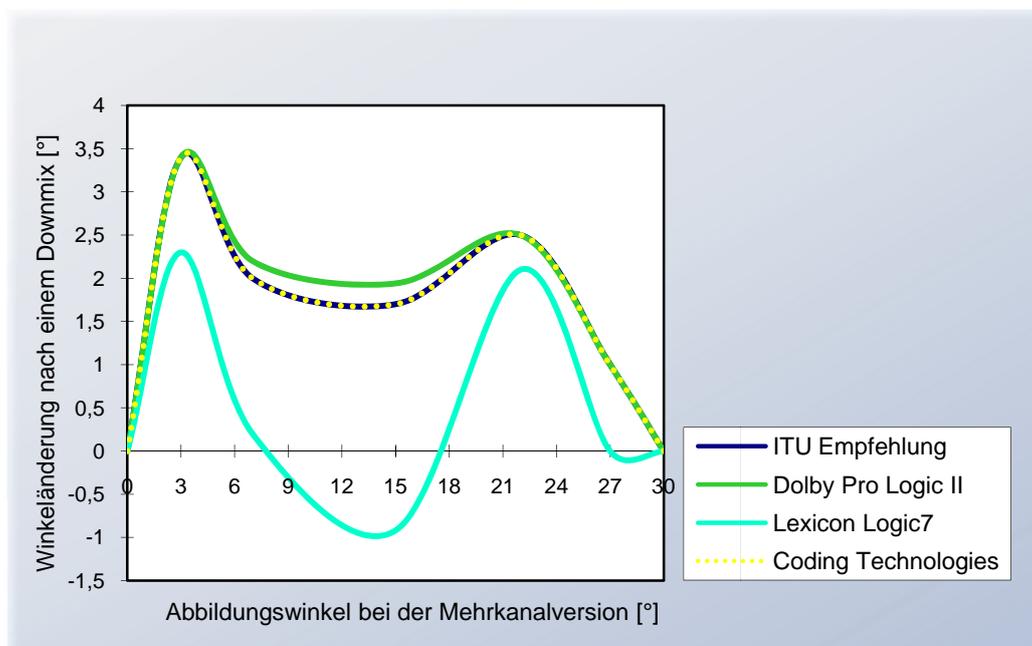
Tabelle 5.4-1 zeigt die von Blauert durch Hörversuche ermittelten Toleranzen bei der Bestimmung des Hörereigniswinkels.

		Knacken	Tonimpulse (Gauß-Töne)				Dauertöne		
			103 Hz	327 Hz	1030 Hz	10,3 kHz	103 Hz	327 Hz	1030 Hz
Lautsprecher- signale mit Pegeldifferenz	0 dB	$\pm 1,3^\circ$	$\pm 2,5^\circ$	$\pm 2^\circ$	$\pm 2^\circ$	$\pm 8^\circ$	$\pm 3,5^\circ$	$\pm 3,5^\circ$	$\pm 5,5^\circ$
	6 dB	$\pm 1,5^\circ$	$\pm 4^\circ$	$\pm 4^\circ$	$\pm 4,5^\circ$	$\pm 6^\circ$	-	-	-
	12 dB	$\pm 3^\circ$	-	-	-	-	$\pm 5^\circ$	$\pm 5^\circ$	$\pm 10^\circ$
	30 dB	$\pm 3,5^\circ$	$\pm 2,5^\circ$	$\pm 2,5^\circ$	$\pm 3,5^\circ$	$\pm 2^\circ$	$\pm 5,5^\circ$	$\pm 5,5^\circ$	$\pm 6^\circ$

**Tabelle 5.4-1** Lokalisationsunschärfe einer durch zwei Lautsprecher erzeugten Phantomschallquelle<sup>87</sup>

Breitbandige Hörereignisse, wie beispielsweise Knackgeräusche, besitzen die besten Lokalisations-eigenschaften. Ein schmalbandiges Signal hingegen, wie zum Beispiel Sprache, wird ungenauer lokalisiert, so dass die um  $1,83^\circ$  verfälschte Darstellung der Phantomschallquelle im ersten Beispiel kaum wahrnehmbar sein dürfte. Die Verschiebung um  $3,4^\circ$ , wie sie im zweiten Beispiel auftritt, kann hingegen durchaus wahrgenommen werden.

In Abbildung 5.4-5 ist die Winkeländerung einer Phantomschallquelle nach verschiedenen Downmixverfahren in Bezug auf ihre Position in der Mehrkanalversion dargestellt. Der Graph eines Downmixes, der keine Winkeländerung hervorruft, würde genau auf der x-Achse liegen.



**Abbildung 5.4-5** Änderung des Abbildungswinkels einer Phantomschallquelle nach einem Downmix in Bezug auf die Mehrkanalversion. Ein positives Vorzeichen entspricht einer Verschiebung nach außen, ein negatives Vorzeichen einer Verschiebung nach innen zur Mitte hin.

<sup>87</sup> vgl. Blauert 1974, S. 167

Unschwer zu erkennen ist, dass die Ergebnisse von Coding Technologies und der ITU eine identische Kurve beschreiben. Coding Technologies basiert auf der ITU-Empfehlung. Der Unterschied zwischen diesen beiden Verfahren liegt hauptsächlich in der Kammfilterkompensation, die Coding Technologies bei der Addition von L und LS bzw. R und RS vornimmt. Die Phantomquellenverschiebung wird, wie im Diagramm ersichtlich, nicht korrigiert. Hingegen führen die Bemühungen von Logic7, die Phantomschallquelle nicht zu verfälschen, zum gewünschten Erfolg: die Verschiebung ist an den Extremstellen mit ca. 2° in einem akzeptablen Bereich.

#### 5.4.2 Phantomschallquellen durch Laufzeitstereofonie

Die Laufzeitstereofonie stellt eine noch größere Herausforderung für Downmixalgorithmen dar. Die Summe der kohärenten, jedoch zeitlich verschobenen Signale erfährt bei der Addition keine über den gesamten Frequenzbereich konstante Pegelanhebung. Abhängig von der Verzögerungszeit werden bestimmte Frequenzen gedämpft oder sogar ausgelöscht, andere wiederum verstärkt. Ein Kammfilter entsteht, der die Klangfarbe verändert. Nach dem Downmix kommen auf Grund der Pegelabsenkung des Centerkanals um 3 dB zusätzliche Pegeldifferenzen zwischen L und R hinzu, die gegenläufig zu den Laufzeitdifferenzen wirken.<sup>88</sup> Die Abbildungseigenschaften der Phantomschallquelle werden dadurch deutlich verschlechtert.

Helmut Wittek und Vincenz Riffeser haben die Abbildung einer durch Laufzeitdifferenz gebildeten Phantomschallquelle zwischen den drei vorderen Lautsprechern nach einem Dolby Pro Logic II Downmix untersucht. Das Ergebnis (Abbildung 5.4-6) zeigt, dass die Lokalisation sehr viel schlechter und das Abbildungsverhalten allgemein enger wird.<sup>89</sup> Auf Grund der gleichen Vorgehensweise bei der Matrizierung der drei vorderen Kanäle von Dolby Pro Logic II und der ITU-Empfehlung ist dieses Ergebnis für beide Verfahren repräsentativ.<sup>90</sup>

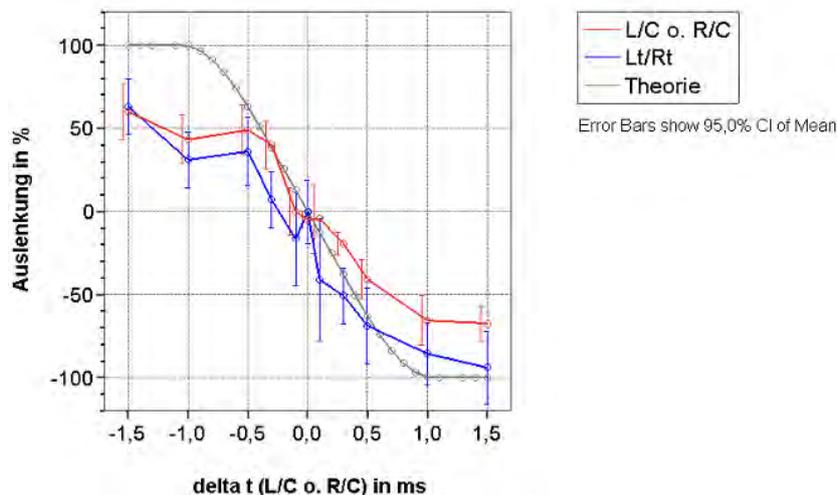


Abbildung 5.4-6 Abbildungsänderung nach einem Dolby Pro Logic II Downmix bei Laufzeitstereofonie

<sup>88</sup> vgl. Deigmöller 2006, S. 37

<sup>89</sup> vgl. Wittek & Riffeser 2005, S. 9

<sup>90</sup> vgl. 3.2 Automatische passive Downmixverfahren

## 6 IRT – Downmix

Nachdem die Schwierigkeiten für einen Downmixalgorithmus im vorhergehenden Kapitel erläutert wurden, steht in diesem Kapitel die Entwicklung des IRT-Downmixes im Vordergrund, die von Martin Keil im Rahmen einer Diplomarbeit vorgenommen wurde.

### 6.1 Wesentliche Ansatzpunkte

Grundlage für die Entwicklung des IRT-Downmixes war das Ergebnis der Evaluation von Jörg Deigmöller. Die Testpersonen der Evaluation waren nach einer Umfrage im Anschluß des Hörversuchs und der Versuchsauswertung der Meinung, dass vorwiegend die Klangfarbe und die Lautstärkebalance bei einem Downmix erhalten bleiben sollten. Als weniger entscheidend wurde die Abbildung bewertet.<sup>91</sup> Entsprechend wurden die Ziele gesteckt: Der neue Downmix sollte vor allem in den Bereichen Klangfarbe und Lautstärkebalance brillieren.

Wie in den vorherigen Kapiteln gezeigt wurde, entsteht ein Großteil der Änderungen in der Klangfarbe durch Kammfiltereffekte. Ein wesentlicher Ansatzpunkt ist daher die Kammfilterkompensation.

Die Ursachen für die Änderung der Lautstärkebalance sind auf die verschiedenen Eigenschaften der Signaladdition zurückzuführen. Additionen die beim Abhören des Mehrkanaltons akustisch geschehen, führen beim Downmix zu einer elektrischen Addition, welche unterschiedliche Ergebnisse für kohärente und inkohärente Signalanteile zur Folge hat.<sup>92</sup> Diese Pegeländerungen im Vergleich zur Mehrkanalversion sollen ausgeglichen werden.

Aber auch bei der Abbildungsstabilität soll der Downmix seiner Konkurrenz nicht nachstehen. Gerade bei Musikaufnahmen mit großen Ensembles soll das auditive Abbild des Downmixes im Frontbereich mit dem der Mehrkanalversion vergleichbar sein.

#### 6.1.1 Kammfilterkompensation

Nach der Untersuchung mehrerer theoretischer Möglichkeiten, einen Kammfiltereffekt zu verhindern oder auszugleichen, entschied sich Keil für die Bearbeitung der Amplituden im Spektralbereich, also der Beträge der DFT-Koeffizienten.

Voraussetzung ist, dass ein Kammfilter detektiert und daraufhin ein Kompensationsvorgang eingeleitet wird, der den Betrag des DFT-Koeffizienten bearbeitet. Optisch gesehen ist das Ziel, der für Kammfilter charakteristischen Wellenform im Spektrum entgegen zu wirken. Hat die Amplitude durch den Kammfilter eine Absenkung erfahren, so wird sie verstärkt, hat sie eine Anhebung erfahren, wird sie gedämpft (Abbildung 6.1-1).

---

<sup>91</sup> vgl. Deigmöller 2006, S. 124

<sup>92</sup> vgl. 4.3 Addition von Amplituden und Pegel

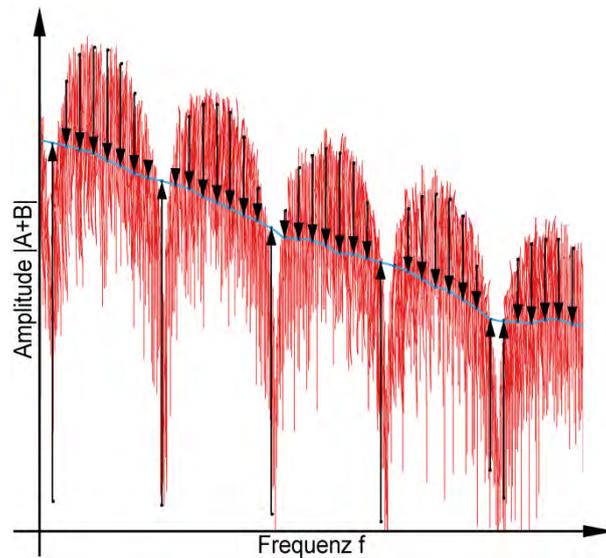


Abbildung 6.1-1 Grafische Darstellung der Veränderung der DFT-Koeffizienten im Spektralbereich zur Kammfilterkompensation.

Um einen Kammfilter detektieren zu können, wird ein Referenzwert für jeden DFT-Koeffizienten benötigt, mit dem dieser verglichen werden kann. Der Referenzwert stellt den Betrag dar, den der DFT-Koeffizient optimaler Weise haben sollte. Für einen solchen Sollwert, der für die Summe zweier Signale  $A[k]$  und  $B[k]$  gilt, bietet sich die energetische Summe an. Diese wird über den Satz des Pythagoras errechnet. In dieser Arbeit wird die Referenz *SOLL* und der Betrag der unbearbeiteten Summe zweier Signale *IST* genannt.

$$IST = |A[k] + B[k]|, \quad (6.1)$$

$$SOLL = \sqrt{|A[k]|^2 + |B[k]|^2}. \quad (6.2)$$

Die energetische Summe zeichnet sich durch einen phasenunabhängigen Betrag aus, der um maximal 3 dB größer ist als der von den zu summierenden Signalen. Sie ist vergleichbar mit der Addition zweier inkohärenter Signale. Dieser Zusammenhang bietet, wie zu sehen sein wird, einige Vorteile und stellt eine sinnvolle Referenz dar.

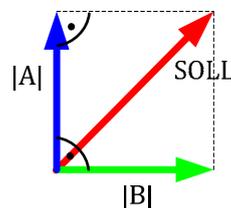


Abbildung 6.1-2 Grafische Darstellung der energetischen Summe  $SOLL = \sqrt{|A[k]|^2 + |B[k]|^2}$ .

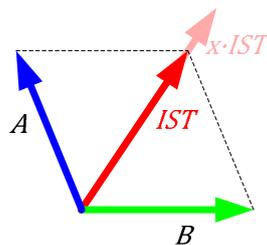
Zur Bearbeitung des DFT-Koeffizienten der Summe wird ein Korrekturfaktor benötigt, mit dem der Koeffizient multipliziert wird. Dieser Korrekturfaktor ist vom Referenzwert und dem Betrag des Koeffizienten abhängig. Soll das Produkt aus Korrekturfaktor  $x$  und *IST*-Wert den *SOLL*-Wert ergeben,

$$x \cdot IST = SOLL \quad (6.3)$$

so ist der Korrekturfaktor  $x$  der Quotient aus  $SOLL$  und  $IST$ .

$$x = \frac{SOLL}{IST}. \quad (6.4)$$

Die Bearbeitung des DFT-Koeffizienten beschränkt sich also auf eine Multiplikation mit dem entsprechenden Korrekturfaktor. Durch diese nachträgliche Manipulation der Amplitude im Spektralbereich ist es theoretisch möglich, den Kammfilter, der durch die Addition zweier Signale entsteht, auszugleichen.

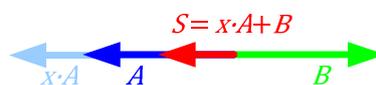


**Abbildung 6.1-3** Grafische Darstellung der Bearbeitung der Summe.

Strebt eine Absenkung bei einem Kammfilter einer Auslöschung entgegen ( $\lim_{IST \rightarrow 0}$ ), so wird der Korrekturfaktor  $x$  unendlich groß.

$$\lim_{IST \rightarrow 0} x = \infty. \quad (6.5)$$

Wird ein DFT-Koeffizient komplett ausgelöscht, so ist es nicht möglich, diesen auf den Referenzwert zu verstärken. In einem solchen Fall ist eine andere Lösung notwendig. Kann die Information des DFT-Koeffizienten nicht mehr aus der Summe gewonnen werden, so bleibt nur die Möglichkeit, auf die beiden zu summierenden Signale zurückzugreifen. Diese müssen bei einer kompletten Auslöschung kohärent und um genau  $180^\circ$  phasenverschoben sein, sowie die gleiche Amplitudengröße aufweisen. Verändert man die Amplitude eines der beiden zu summierenden DFT-Koeffizienten, verstärkt oder dämpft diese also, so kann die vollständige Auslöschung verhindert werden.



**Abbildung 6.1-4** Grafische Darstellung der Bearbeitung einer der beiden zu summierenden DFT-Koeffizienten, so dass eine Auslöschung in der Summe verhindert wird.

Dieser Umstand bewog Keil, DFT-Koeffizienten, deren Summe 60% oder weniger des SOLL-Werts aufweisen, vor der Addition zu bearbeiten.

### 6.1.2 Erhaltung der effektiven Energie

Wie in Kapitel 5.2 gezeigt wurde, verändert ein Downmix den resultierenden Schalldruckpegel an der Abhörposition. Ursache sind die unterschiedlichen Additionseigenschaften kohärenter und inkohärenter sowie akustischer und elektrischer Signale.<sup>93</sup>

Durch die Korrektur der DFT-Koeffizienten mit einem Referenzwert, der 3 dB größer ist als die zu summierenden Signale, wird bei der Kammfilterkompensation auch die bei der elektrischen Addition von kohärenten Signalen auftretende Erhöhung des Pegels um bis zu 6 dB korrigiert, die für die Änderung des Schalldruckpegels beim Hörer nach dem Downmix verantwortlich ist. Die Vorgehensweise bei der Kammfilterkompensation sollte damit auch die Erhaltung der effektiven Energie sichern.

### 6.1.3 Differenzpegeländerung kohärenter und inkohärenter Signalanteile

Ziel ist es, den Differenzpegel, der beim Abhören des Mehrkanaltons an der Position des Sweet-Spots entsteht, auch bei der herunter gemischten Zweikanalversion zu erhalten. Durch eine Bearbeitung des Summensignals, die sich an der energetischen Summe orientiert - also an einer maximalen Pegelzunahme von 3 dB pro Addition - wird auch die Differenzpegeländerung unterdrückt.

### 6.1.4 Phantomschallquellenverschiebung

Bei einem Downmix ist eine Verschiebung der Phantomschallquellen, die mit Hilfe des Center-Lautsprechers gebildet werden, von der Medianebene<sup>94</sup> nach außen hin zu verzeichnen.<sup>95</sup> Die Ursache liegt auch hier in der um 3 dB stärkeren Pegelzunahme bei der elektrischen Addition des kohärenten und phasengleichen Signals aus dem Center- und einem der beiden seitlichen Frontlautsprecher (L oder R). Mit der Bearbeitung der DFT-Koeffizienten im Rahmen der Kammfilterkompensation, welche die Amplitude des Koeffizienten der Summe bei einer Pegelzunahme von mehr als 3 dB korrigiert, ist auch eine Kompensation der Phantomschallquellenverschiebung zu erwarten.

### 6.1.5 Eine Lösung für vier Probleme

Das Konzept der Kammfilterkompensation mit einer Korrektur der DFT-Koeffizienten, die sich an der energetischen Summe orientiert, verhindert oder entschärft theoretisch alle vier wesentlichen Probleme, welche bei einem Downmix auftreten. Der Grund dafür ist in der Tatsache zu finden, dass drei der vier Probleme auf die verschiedenen Additionseigenschaften von kohärenten, inkohärenten, elektrischen und akustischen Signalen zurückzuführen sind.

## 6.2 Realisierung des automatischen Downmixes

Gegenstand dieser Diplomarbeit ist neben der qualitativen Optimierung des Downmixalgorithmus auch die Vereinfachung der Implementierung, also eine quantitative Optimierung. Diese trägt wesentlich zur Überschaubarkeit des Quelltextes und der Performance<sup>96</sup> des Programms bei. Es wäre müßig, an dieser Stelle von Keils Quelltext auszugehen, um später die quantitative Opti-

---

<sup>93</sup> vgl. 4.3 Addition von Amplituden und Pegel

<sup>94</sup> vgl. 4.6 Kopfbezogenes Koordinatensystem

<sup>95</sup> vgl. 5.4 Phantomschallquellenverschiebung

<sup>96</sup> Die Performance bezeichnet die Leistungsfähigkeit eines Programms.

mierung zu beschreiben. Daher bedient sich dieses Kapitel bereits der optimierten Form der Implementierung auf dem qualitativen Stand des Algorithmus von Martin Keil.<sup>97</sup> Keil's Variante des IRT-Downmix wird in der vorliegenden Arbeit mit dem Kürzel „IRT\_5keil“ bezeichnet.

### 6.2.1 Ablauf des Downmixvorgangs

Der Downmixvorgang wird in vier Additionsstufen unterteilt, in denen die fünf Signale des Mehrkanalton zu den zwei Kanälen des Downmixes gemischt werden (Abbildung 6.2-1). Als erstes erfolgt die Mischung der drei Frontkanäle.

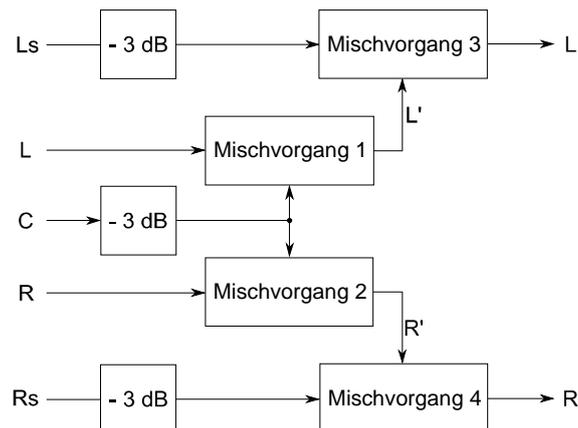


Abbildung 6.2-1 Blockschaltbild IRT-Downmix

In den Stufen 1 und 2 wird der um 3 dB abgesenkte Centerkanal mit jeweils dem linken und dem rechten Frontkanal addiert. Die beiden daraus entstehenden Mischsignale werden anschließend in den Stufen 3 und 4 mit den seitengleichen, jeweils um 3 dB abgesenkten Surroundkanälen addiert. Die daraus entstandenen Mischsignale stellen den linken bzw. rechten Kanal des Stereo-Downmixes dar.<sup>98</sup>

Das Schaubild in Abbildung 6.2-3 veranschaulicht den Ablauf des Mischvorgangs, wie er in allen vier Stufen umgesetzt wird.

Um ein kontinuierliches Signal analytisch verarbeiten zu können, ist es notwendig, dieses in Zeitfenster zu unterteilen (Abbildung 6.2-3 (1)). Diese Fenster werden dann sequentiell verarbeitet und anschließend wieder zu einem ganzen Signal zusammengesetzt.<sup>99</sup> Um keine abrupten Übergänge zwischen den wieder zusammengesetzten Fenstern zu erhalten und um den Leakage-Effect<sup>100</sup> zu unterdrücken, werden diese mit 50% Überlappung aus dem zu verarbeitenden Signal

<sup>97</sup> Für den originalen Quelltext von Martin Keil vgl. Keil 2007, S. 65.

<sup>98</sup> Die Absenkung des Centerkanals und der Surroundkanäle vor dem Additionsprozess soll später je nach Genre variabel einstellbar sein. In dieser Arbeit wird jedoch der Einfachheit halber mit der Standard-Dämpfung von -3 dB gearbeitet.

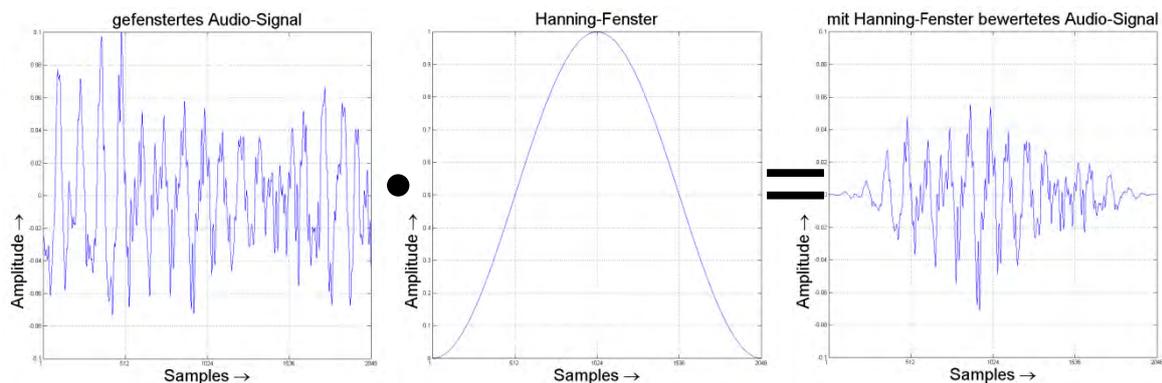
<sup>99</sup> vgl. 4.8.1 Vom Zeit- in den Frequenzbereich

<sup>100</sup> Der Leakage Effect - auch Leck-Effekt genannt - bezeichnet das Auftreten von Frequenzen im diskreten Spektrum, die nicht im Ausgangssignal enthalten sind. Diese entstehen, wenn die Fensterlänge nicht der Periodendauer aller im Signal enthaltenen Frequenzen oder deren ganzzahligen Vielfachen entspricht. Vgl. Werner 2006, S. 59.

ausgeschnitten und mit der Hann-Funktion<sup>101</sup> bewertet (Abbildung 6.2-2). Die Bewertung des Fensters  $a_{\text{Fenster}}[n]$  wird durch dessen Multiplikation mit einem Hanning-Fenster  $w[n]$  gleicher Länge realisiert und bewirkt beim Zusammensetzen der Fenster einen weichen Übergang ähnlich wie bei einem Crossfade<sup>102</sup>. Zur Erstellung des Hanning-Fensters  $w[n]$  dient die Hann-Funktion, wobei  $M$  die verwendete Fensterlänge ist:

$$w(n) = \frac{1}{2} \left[ 1 + \cos \left( \frac{2\pi n}{M} \right) \right], \quad n = -\frac{M}{2}, \dots, \frac{M}{2}; \quad (6.6)$$

$$a_{\widehat{\text{Fenster}}}[n] = a_{\text{Fenster}}[n] \cdot w[n]. \quad (6.7)$$



**Abbildung 6.2-2** Bewertung eines gefenstersten Audiosignals mit einem Hanning-Fenster.

Die Multiplikation eines gefenstersten Signals mit einer das Fenster bewertenden Folge wird auch Hüllkurvenbewertung genannt (Abbildung 6.2-3 (2)).

Martin Keil entschied sich für eine Fenstergröße von 2048 Abtastwerten (engl. samples). Um Aliasing-Artefakte zu vermeiden<sup>103</sup>, werden den 2048 Abtastwerten des Eingangssignals noch 2048 Nullen angehängt, so dass der neu entstandene Block aus 4096 Abtastwerten besteht. Das Auffüllen eines Fensters mit Nullen nennt man Zero-Padding (Abbildung 6.2-3 (3)).<sup>104</sup>

Anschließend wird dieses 4096 Abtastwerte große Fenster mit Hilfe einer FFT in den Frequenzbereich (auch Spektralbereich oder Frequenzspektrum genannt) überführt. Dieser Frequenzbereich besteht ebenfalls aus 4096 Werten, den sogenannten komplexen DFT-Koeffizienten. Jeder DFT-Koeffizient enthält für die ihm zugeteilte Harmonische (Frequenz) Informationen über die Amplitudengröße sowie die Phasenverschiebung.<sup>105</sup>

<sup>101</sup> Funktion, die zur Erstellung des Hanning-Fensters verwendet wird (vgl. (6.6)). Die Bezeichnung rührt aus der Publikation „Particular Pairs of Window.“ von R. B. Blackman und J. W. Tukey (vgl. Blackman & Tukey 1959, S. 98 f.). Blackman und Tukey benannten die Funktion nach Julius von Hann. Die Bezeichnung „Hanning-Fenster“ entstammt auch diesem Artikel, bezieht sich jedoch auf die Anwendung des Fensters, abgeleitet von „to hann“.

<sup>102</sup> Unter einem Crossfade versteht man einen kontinuierlichen Übergang von einem Signal zum anderen. Die voneinander abhängigen Dämpfungen der beiden Signale folgen dabei einer Funktion, die z.B. linear oder logarithmisch sein kann. Ziel der Funktionen ist zum Beispiel die Erhaltung der effektiven Energie oder der Lautstärke.

<sup>103</sup> vgl. Keil 2007, S. 35

<sup>104</sup> vgl. Werner 2006, S. 58 f.

<sup>105</sup> vgl. 4.8.1 Vom Zeit- in den Frequenzbereich

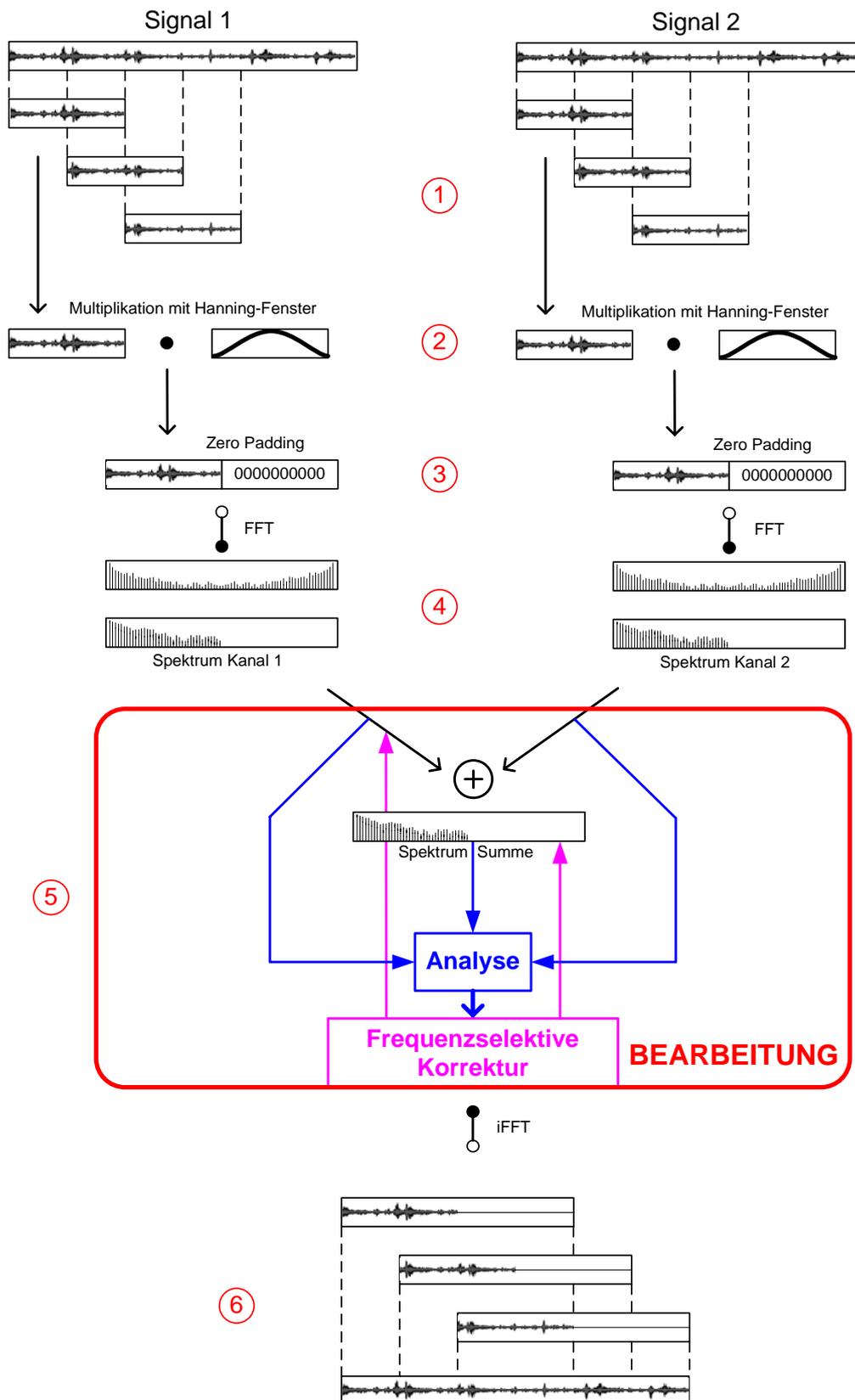


Abbildung 6.2-3 Blockdarstellung des IRT-Downmixverfahrens nach Keil.

Auf Grund der Redundanz im achsensymmetrischen Frequenzbereich reicht es aus, nur eine Hälfte des Spektrums zu bearbeiten. Die zweite Hälfte, auch Spiegelspektrum genannt, wird

daher auf Null gesetzt. Um die Energie des Signals zu erhalten, werden die Amplituden der Koeffizienten in der ersten Hälfte verdoppelt. Diese Verdopplung geschieht mit einer Ausnahme, dem ersten DFT-Koeffizienten. Dieser Koeffizient liegt auf der Spiegelungsachse und ist für den Gleichspannungsanteil im Signal verantwortlich (Abbildung 6.2-3 (4)).

Im Spektralbereich findet die eigentliche Addition der beiden Signale mit anschließender Bearbeitung des Summenspektrums statt. Dazu werden die beiden im Spektralbereich vorliegenden, zu summierenden Signale addiert. Natürlich wäre auch eine Addition im Zeitbereich möglich, diese hätte jedoch auf Grund von drei FFTs eine Verschlechterung der Performance zur Folge (Abbildung 6.2-3 (5)).

Nach der Addition und der Bearbeitung folgt die inverse FFT (iFFT) und anschließend das Zusammensetzen der einzelnen Fenster. Wegen des Zero-Paddings erfolgt dies mit einer  $\frac{3}{4}$ -Überlappung, das heißt, der Bereich mit den Nullen wird einfach mit eingearbeitet. Das sich daraus ergebende Signal ist die korrigierte Mischung der beiden zu summierenden Signale (Abbildung 6.2-3 (6)).

## 6.2.2 Der Algorithmus

Im Spektralbereich stehen zur weiteren Bearbeitung die beiden zu summierenden Signale sowie deren Summe zur Verfügung. Die Summe entspricht dem IST-Zustand nach der Addition. Der Betrag eines DFT-Koeffizienten der Summe entspricht somit dem IST-Wert für die dem Koeffizienten zugewiesene Frequenz.

Zur Detektion eines Kammfilters wird mit Hilfe von (6.2) die energetische Summe der beiden Eingangssignale ermittelt und als Referenz verwendet, so dass für jeden IST-Wert ein SOLL-Wert existiert, mit dem er verglichen werden kann.<sup>106</sup> Aus diesem Vergleich folgen eine Fallunterscheidung und die dementsprechende Korrektur des DFT-Koeffizienten.

### Fall 1: Der IST-Wert ist kleiner als 10% des SOLL-Werts

Wenn der *IST*-Wert, also der Betrag des DFT-Koeffizienten, kleiner als 10% des *SOLL*-Werts ist, dann löschen sich die beiden zu summierenden Signale annähernd oder vollständig aus. Damit liegt ein Notch eines Kammfilters vor, bei der die Phasendifferenz der beiden Eingangssignale genau oder nahezu  $180^\circ$  beträgt.

Um solch eine Auslöschung zu verhindern, verstärkt Keil das erste der beiden Eingangssignale generell um 190%, um danach die Addition erneut durchzuführen.

$$|A_{\text{bearbeitet}}[k]| = |A[k]| \cdot 1,9, \quad (6.8)$$

$$S_{\text{bearbeitet}}[k] = A_{\text{bearbeitet}}[k] + B[k]. \quad (6.9)$$

<sup>106</sup> vgl. 6.1.1 Kammfilterkompensation

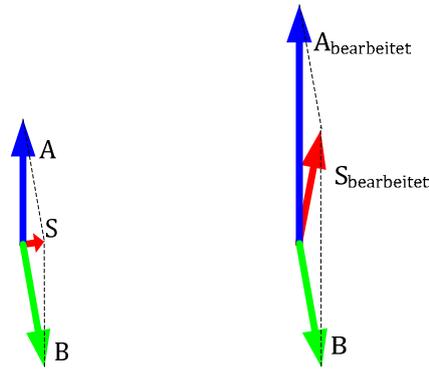


Abbildung 6.2-4 Darstellung der Bearbeitung des IRT-Downmixes in der komplexen Ebene bei Fall 1.

### Fall 2: Der *IST*-Wert liegt zwischen 10% und 60% des *SOLL*-Werts

Wie in Fall 1 liegt hier eine zu starke Dämpfung der Amplitude vor. Ursache sind Phasenwinkel der beiden zu summierenden Signale, die zwar keine vollständige Auslöschung zur Folge haben, aber eine spürbare Absenkung.

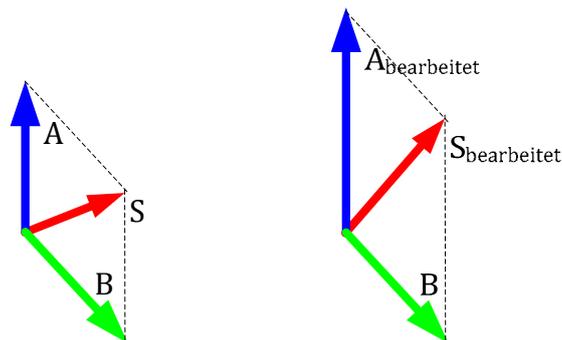


Abbildung 6.2-5 Darstellung der Bearbeitung des IRT-Downmixes in der komplexen Ebene bei Fall 2.

Um eine solche Auswirkung zu reduzieren, erfolgt eine Anhebung des Amplitudenwertes des ersten Eingangssignals in Abhängigkeit zur Stärke des Einbruchs im Summensignal. Keil wählte eine lineare Verstärkung mit einem Faktor zwischen 1,4 und etwa 1,9.

$$|A_{\text{bearbeitet}}[k]| = |A[k]| \cdot \left( \frac{\text{SOLL}}{\text{IST}} \cdot 0,05 + 1,4 \right), \quad (6.10)$$

$$S_{\text{bearbeitet}}[k] = A_{\text{bearbeitet}}[k] + B[k]. \quad (6.11)$$

### Fall 3: Der *IST*-Wert ist etwa gleich groß wie der *SOLL*-Wert (60%-100% des *SOLL*-Werts)

Tritt dieser Fall ein, so liegt der *IST*-Wert in einem akzeptablen Bereich (60%-100% des *SOLL*-Werts) und muss nach Ansicht von Keil nicht korrigiert werden. Man kann davon ausgehen, dass in diesem Bereich vor allem inkohärente Frequenzanteile miteinander addiert werden und weder eine Peak- noch eine Notch-Frequenz eines Kammfilters vorliegt.

#### Fall 4: Der IST-Wert ist größer als der SOLL-Wert

Dieser Fall tritt ein, wenn die Differenz der Phasenwinkel zwischen den beiden zu summierenden Signalen kleiner als  $90^\circ$  ist. Der Grund hierfür kann sowohl eine Addition zweier kohärenter und phasengleicher Signale als auch das Auftreten eines Maximums bei einem Kammfilter sein.

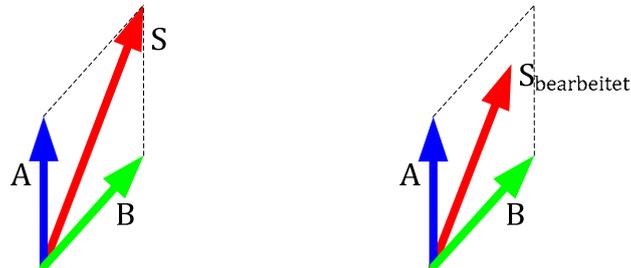


Abbildung 6.2-6 Darstellung der Bearbeitung des IRT-Downmixes in der komplexen Ebene bei Fall 4.

Bei der Korrektur der Amplitude wird der zu starke Anstieg auf 10% reduziert. Zur Berechnung dient die folgende Formel:

$$|S_{\text{bearbeitet}}[k]| = SOLL + (IST - SOLL) \cdot 0,1. \quad (6.12)$$

#### 6.2.3 Struktur der Implementierung

Dieses Kapitel führt in die bereits vom Autor der vorliegenden Arbeit optimierte Implementierung des Algorithmus ein. Verwendet wurde MATLAB Version 7.3.0.267 (R2006b) mit der Erweiterung „*Signal Processing Toolbox*“.

Da der Algorithmus zu einem großen Teil aus wiederkehrenden Bearbeitungsschritten besteht, bietet sich bei der Implementierung die Verwendung sogenannter Funktionen an.<sup>107</sup> Abbildung 6.2-7 zeigt die daraus resultierende modulare Aufbauform des Programms.

<sup>107</sup> Eine Funktion bezeichnet in der Informatik ein Unterprogramm, das über den Namen der Funktion immer wieder aufgerufen werden kann. Beim Aufrufen der Funktion können ein oder mehrere Parameter übergeben werden. Ebenso kann die Funktion am Ende einen Funktionswert zurückgeben. Damit sind Funktionen in der Informatik ein wichtiges Mittel zur Modularisierung. Vgl. Schweizer 2007, S. 87.

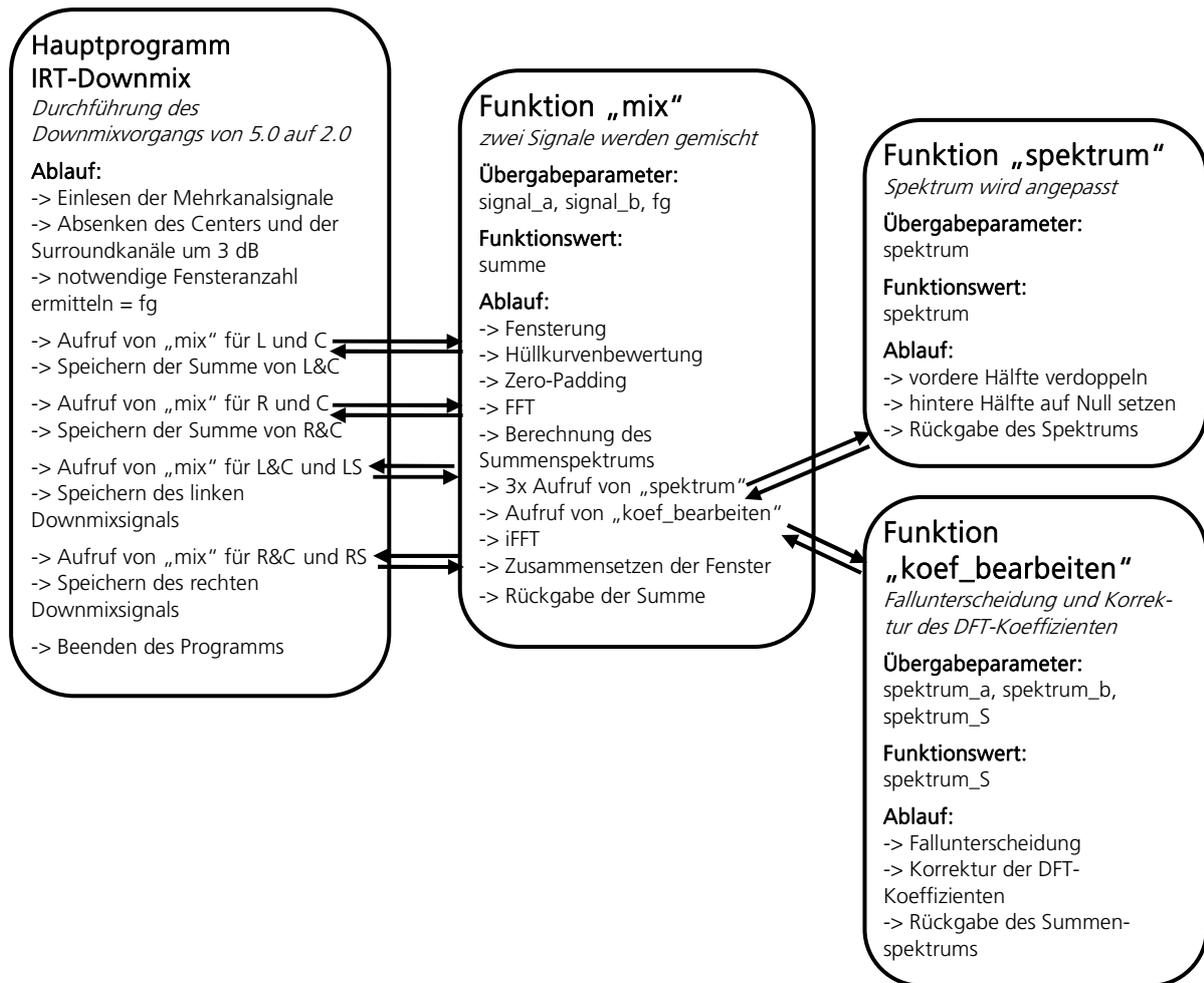


Abbildung 6.2-7 Modulare Darstellung des IRT-Downmixes

Das Programm besteht aus einem Hauptprogramm und drei Funktionen, auf welche das Hauptprogramm direkt oder indirekt über die Funktion „mix“ zugreifen kann. Das Hauptprogramm liest die wav-Dateien des Mehrkanaltons ein, senkt den Centerkanal sowie die Surroundkanäle um 3 dB ab und ermittelt die notwendige Anzahl von Fenstern, in welche die Eingangssignale unterteilt werden müssen. Danach wird für alle Signalpaare, die addiert werden sollen, die Funktion „mix“ aufgerufen, die das gemischte Signal zurückgibt. Dieses muss entweder - wenn es sich um eine vorläufige Mischung, wie die von L und C handelt - zwischengespeichert oder - wenn es sich um eines der beiden Ausgangssignale handelt – endgültig abgespeichert werden. (vgl. Quelltext: *IRT-Downmix, Version „IRT\_5keil“ S. 103; Zeilen 15-117*)

Erst die Funktion „mix“, welche die beiden zu summierenden Signale sowie die notwendige Fensteranzahl erhalten hat, nimmt in einer Schleife die Fensterung beider Signale vor, bewertet jedes Fenster mit der Hanning-Funktion, führt das Zero-Padding durch und überführt jedes Fenster per FFT in den Spektralbereich. Dort werden die beiden zugehörigen Fenster addiert, so dass man das Spektrum der Summe erhält.

Alle Spektren werden mit der Funktion „spektrum“ angepasst und anschließend der Funktion „koef\_bearbeiten“ übergeben, wo sie mit dem Ziel der Kammfilterkompensation bearbeitet werden. Der Funktionswert, der von der Funktion „koef\_bearbeiten“ zurückgegeben wird, ist

das korrigierte Summenspektrum, das mit Hilfe der iFFT wieder in den Zeitbereich transformiert wird. Dort werden die einzelnen Fenster wieder zusammengesetzt und die Mischung an das Hauptprogramm zurückgegeben. (vgl. *Quelltext: IRT-Downmix, Version „IRT\_5keil“ S. 103; Zeilen 150-225*)

Die Funktion „spektrum“ bereitet das ihr übergebene Spektrum für die weitere Bearbeitung vor. Dazu werden die Werte der Amplituden, mit Ausnahme des Gleichanteils, in der ersten Hälfte verdoppelt. Die zweite Hälfte des Spektrums wird auf Null gesetzt.<sup>108</sup> Das auf diese Weise angepasste Spektrum wird wieder an die Funktion „mix“ zurückgegeben. (vgl. *Quelltext: IRT-Downmix, Version „IRT\_5keil“ S. 103; Zeilen 228-247*)

Der eigentliche Kern des Programms - der Algorithmus - findet sich in der Funktion „koef\_bearbeiten“ wieder. Der Funktion werden sowohl das Spektrum der Summe als auch die zwei Spektren der zu summierenden Signale übergeben. Aus diesen beiden Signalen werden die SOLL-Werte über die energetische Summe berechnet. In einer „if-else“-Schleife werden die DFT-Koeffizienten des Summenspektrums, also die IST-Werte, mit ihren zugehörigen SOLL-Werten verglichen und gemäß Kapitel 6.2.2 korrigiert. Das fertige Summenspektrum wird zum Schluss an die Funktion „mix“ zurückgegeben. (vgl. *Quelltext: IRT-Downmix, Version „IRT\_5keil“ S. 103; Zeilen 251-292*)

---

<sup>108</sup> vgl. Kapitel 6.2.1 Ablauf des Downmixvorgangs

## 7 Analyse und Optimierung des bisherigen Downmixes hinsichtlich der Kammfilterkompensation

Das Wort „Optimierung“ stammt von dem lateinischen Wort „optimum“ ab, was übersetzt „das Beste“ bedeutet. Eine Optimierung soll bei einem Vorgang oder Zustand bezüglich eines oder mehrerer Gesichtspunkte Schwachstellen beseitigen, das Ergebnis verbessern und/oder die Effizienz steigern.<sup>109</sup> Im Falle dieser Arbeit handelt es sich bei dem Vorgang um einen automatischen Stereo-Downmix und bei den zu betrachtenden Gesichtspunkten um die subjektive wie objektive Qualität.

Um eine Optimierung realisieren zu können, darf das Optimum des Vorgangs noch nicht erreicht sein. Vor der Optimierung steht daher eine Analyse, welche die Schwachstellen und Probleme aufdeckt.

### 7.1 Subjektiver Eindruck des IRT-Downmix

Werden mit dem Downmix „IRT\_5keil“ erstellte Stereoverionen unterschiedlicher Genres auf Schwachstellen und Probleme hin angehört, so fallen, neben dem in Kapitel 3.4.2 erwähnten guten Ergebnis hinsichtlich der Kammfilterkompensation, zwei Dinge auf:

Zum einen können bei besonders tonalen Signalen Artefakte wahrgenommen werden, die sich in einer an- und abschwellenden Rauigkeit des Klangs bis hin zu deutlich hörbaren Knackgeräuschen äußern.<sup>110</sup>

Zum anderen erscheinen das klangliche Abbild und die Räumlichkeit des Downmixes im Vergleich zur Mehrkanalversion im Frontbereich gelegentlich unstimmig.

### 7.2 Optimierung durch Analysen im Zeitbereich

Um die subjektiv wahrgenommenen Artefakte näher definieren zu können, werden objektive Analysen mit einer Auswertung im Zeitbereich vorgenommen.

#### 7.2.1 Grafische Darstellung der Bearbeitung im Zeitbereich

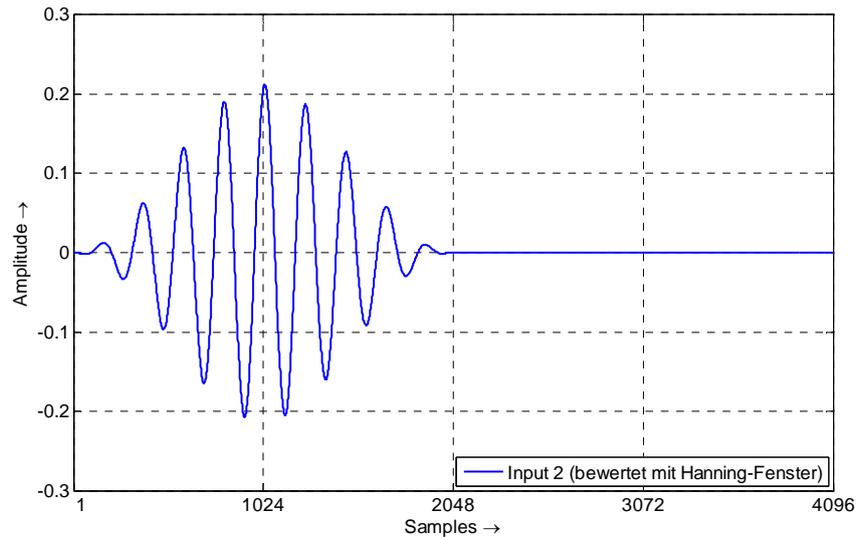
Zur grafischen Beurteilung der Bearbeitung im Zeitbereich wird ein Testsignal zu verschiedenen Zeitpunkten während des Downmixes grafisch dargestellt. Die Auswertung dieser Grafiken ermöglicht Rückschlüsse auf die verschiedenen Abschnitte des Downmixes und dadurch eine gezielte Fehlersuche.

Beispielhaft für alle vier Additionsvorgänge wird für die Analyse - sofern nicht anders beschrieben - der Additionsprozess von L und C herangezogen. Zu beachten ist hierbei die Dämpfung des Centerkanals, die während des Downmixvorgangs vorgenommen wird.

Für den zeitlichen Darstellungsbereich bietet sich die Größe eines Fensters inklusive dem Zero-Padding an, also 4096 Samples. Ein Fenster nach der Hüllkurvenbewertung und dem Zero-Padding ist in der Abbildung 7.2-1 abgebildet.

<sup>109</sup> vgl. Brockhaus 1968, S. 68

<sup>110</sup> vgl. Hörbeispiel 1: Chor auf Audio-DVD



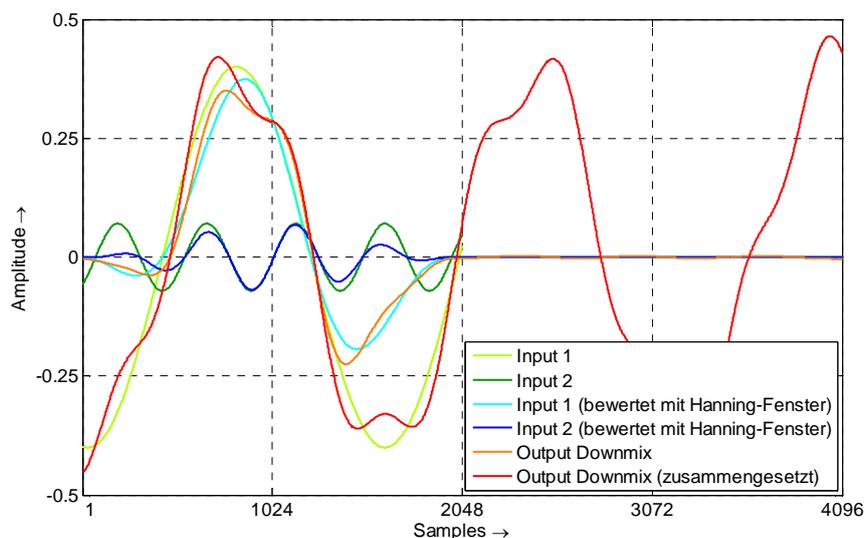
**Abbildung 7.2-1** Fenster aus dem Input-Signal  $c[n] = \sin(2\pi \cdot 200\text{Hz} \cdot nT)$  mit einer Abtastfrequenz von  $f_s = 44,1 \text{ kHz}$  nach der Hüllkurvenbewertung

Man kann im ersten Teil (bis zum 2048. Sample) deutlich das gefensterte und mit der Hanning-Funktion bewertete Sinussignal erkennen. Im zweiten Teil liegen die durch das Zero-Padding erzeugten Werte - wie erwartet - konstant bei Null.

Insgesamt werden sechs „Signalproben“ während des Downmixvorgangs entnommen:

- Die beiden Eingangssignale nach der Fensterung.
- Die beiden Eingangssignale nach der Hüllkurvenbewertung und dem Zero-Padding.
- Das bearbeitete und noch gefensterte Summensignal (Output) nach der Rücktransformation aus dem Spektralbereich.
- Der zugehörige Bereich aus dem wieder zusammengesetzten Signal.

Legt man diese sechs Signale grafisch übereinander, so erhält man Abbildung 7.2-2.



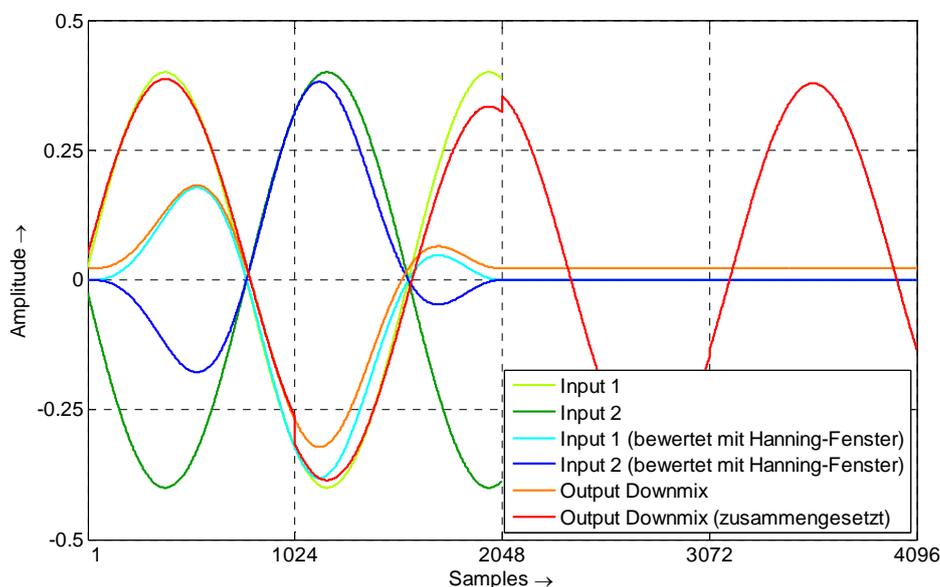
**Abbildung 7.2-2** Fenster mit allen sechs Signalproben. Ergebnis der Eingangssignale  $l[n] = \sin(2\pi \cdot 30\text{Hz} \cdot nT)$  und  $c[n] = \sin(2\pi \cdot 100\text{Hz} \cdot nT)$  mit einer Abtastfrequenz von  $f_s = 44,1 \text{ kHz}$ .

Hell- und dunkelgrün sind die beiden noch unbewerteten Eingangssignale dargestellt. Multipliziert mit dem Hanning-Fenster ergeben diese beiden Eingangssignale den hell- und dunkelblauen Graphen. Da die Hanning-Funktion nur genau in der Mitte den Wert 1 hat, in diesem Beispiel also beim 1024. Sample, hat das bewertete Signal auch nur an dieser Stelle seine ursprüngliche Größe der Amplitude. Die Graphen des bewerteten und des unbewerteten Signals berühren sich daher lediglich in diesem Punkt. Das Ergebnis des Downmixes für das dargestellte Fenster zeigt der orangefarbene Graph. Wird das Output-Signal dieses Fensters mit den Signalen der vorhergehenden und nachfolgenden Fenster zusammengesetzt, so erhält man das kontinuierliche Output-Signal, welches der rote Graph veranschaulicht.

### 7.2.2 Analyse mit synthetischen Signalen

Die erste Analyse soll mit einem synthetisch generierten Signal durchgeführt werden. Anhand solcher vorhersehbarer Signale kann eine Prognose aufgestellt werden, wie der Downmix aussehen müsste.

Um zu sehen, ob die Kammfilterkompensation richtig arbeitet, werden für L und C zwei Sinussignale mit der gleichen Frequenz aber entgegengesetztem Phasengang erzeugt. Das Center-Signal wird vor dem Downmixvorgang um 3 dB verstärkt, so dass die inbegriffene Dämpfung um 3 dB kompensiert und eine vollständige Auslöschung erreicht wird. Der Downmix sollte ein Signal ausgeben, das - im Gegensatz zur einfachen Addition der beiden Signale - nicht völlig ausgelöscht ist.



**Abbildung 7.2-3** Fenster der beiden Signale  $I[n] = 0,4 \cdot \sin(2\pi \cdot 30\text{Hz} \cdot nT)$  und  $c[n] = 0,4 \cdot \sqrt{2} \cdot \sin(2\pi \cdot 30\text{Hz} \cdot nT + \pi)$  aus dem IRT-Downmix nach Martin Keil (IRT\_5keil). Die Bearbeitung der DFT-Koeffizienten beginnt mit dem zweiten Koeffizienten.

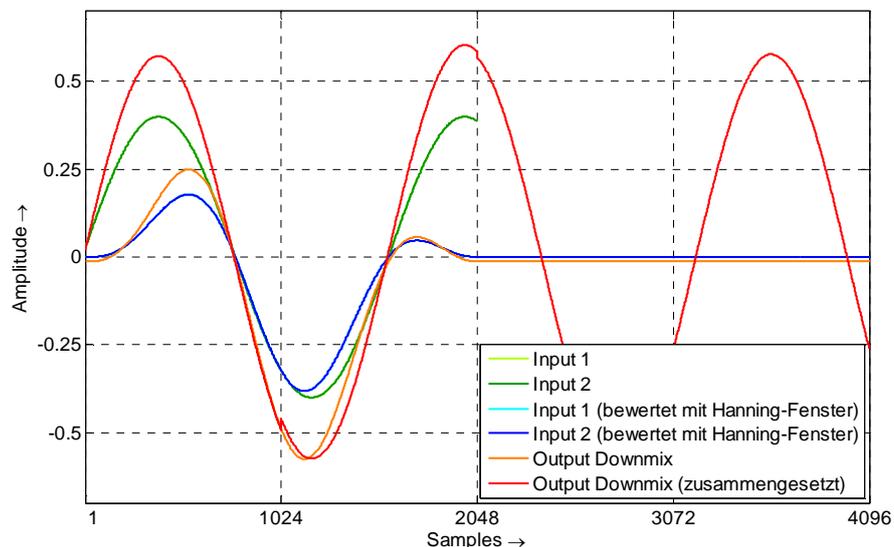
Wie erwartet kompensiert der Algorithmus die Auslöschung und gibt ein Signal mit der gleichen Frequenz und einer Amplitude aus, die etwa so groß ist wie die der beiden Eingangssignale. Der Kammfilter wird also nicht komplett korrigiert, ist doch die energetische Summe zwei gleich großer Amplituden um etwa 3 dB größer. Der Grund hierfür liegt in der Bearbeitungsmethodik.

Im Fall 1 der Fallunterscheidung, der für diese Signalkonstellation zutrifft, wird der IST-Wert bewusst nicht vollständig auf den SOLL-Wert korrigiert.<sup>111</sup>

Auffällig hingegen sind sowohl die Amplitudensprünge des zusammengesetzten Output-Signals bei den Samples 1024, 2048 und 3072 als auch die deutlich erkennbare Anhebung aus der x-Achse des orangefarbenen Output-Signals im zweiten Teil des Diagramms. Betrachtet man den Anfang des Fensters, so entdeckt man, dass auch dort das Output-Signal nicht bei Null, sondern im positiven Bereich beginnt. Diese Merkmale deuten auf einen Gleichspannungsanteil hin, der beim Downmix entstanden ist.

Die Untersuchung weiterer Fenster ergab, dass dieser Gleichspannungsanteil von Fenster zu Fenster variiert und zudem auch negative Werte annehmen kann. Werden die Fenster wieder zusammengesetzt, entstehen an den 1024 Samples auseinanderliegenden Schnittstellen die im roten Graphen deutlich erkennbaren Signalsprünge.

Ein zweiter Versuch soll den Bearbeitungsfall 4 provozieren, in dem der Algorithmus auf eine zu starke Anhebung des Summensignals im Vergleich zum SOLL-Wert reagiert. Dazu wird das Center-Signal aus dem vorangegangenen Versuch um  $180^\circ$  in der Phase gedreht, so dass es zur Addition zweier frequenz- und pegelgleicher Signale ohne Phasenverschiebung kommt. Das Ergebnis in Abbildung 7.2-4<sup>112</sup> lässt ebenfalls auf eine korrekte Kompensation des Kammfilters schließen. Das Output-Signal weist eine Amplitude von 0,6 anstelle von 0,8 auf und korrigiert damit das Peak gewollt nicht vollständig, aber nahezu. Jedoch sind auch in diesem Versuch deutlich der Gleichspannungsanteil beim orangefarbenen Output-Signal und die Signalsprünge beim roten Graphen des zusammengesetzten Output-Signals zu erkennen. Der Fehler ist also nicht abhängig von der fallweisen Art der Bearbeitung.



**Abbildung 7.2-4** Fenster der beiden Signale  $l[n] = 0,4 \cdot \sin(2\pi \cdot 30\text{Hz} \cdot nT)$  und  $c[n] = 0,4 \cdot \sqrt{2} \cdot \sin(2\pi \cdot 30\text{Hz} \cdot nT)$  aus dem IRT-Downmix nach Martin Keil (IRT\_5keil). Die Bearbeitung der DFT-Koeffizienten beginnt mit dem zweiten Koeffizienten.

<sup>111</sup> vgl. Kapitel 6.2.2 Der Algorithmus

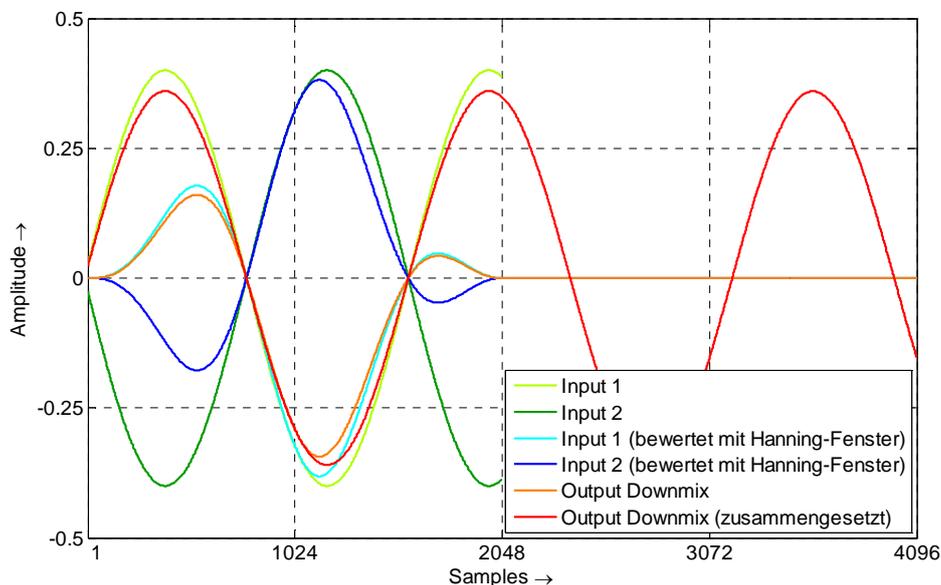
<sup>112</sup> Da die Signale  $l[n]$  und  $c[n]$  identisch sind, wird bei der grafischen Darstellung das Input-Signal 1 vom Input-Signal 2 verdeckt.

### 7.2.3 Gleichspannungsanteil nach dem Additionsvorgang

Die Ursache für den im Downmixprozess entstehenden Gleichspannungsanteil ist in der Bearbeitung des ersten DFT-Koeffizienten zu suchen. Es gibt genau zwei Stellen im gesamten Vorgang, an denen DFT-Koeffizienten bearbeitet bzw. verändert werden: Einmal in der Funktion „spektrum“, in welcher das Spektrum für die Bearbeitung angepasst wird. Zum zweiten in der Funktion „koef\_bearbeiten“, in der die DFT-Koeffizienten hinsichtlich der Kammfilterkompensation bearbeitet werden.

Betrachtet man Keils Implementierung genauer, so fällt auf, dass weder in der Funktion „spektrum“ noch in der Funktion „koef\_bearbeiten“ der erste DFT-Koeffizient tatsächlich verändert wird. Beide Funktionen beginnen ihre Bearbeitung erst mit dem zweiten Koeffizienten (vgl. *Quelltext: IRT-Downmix, Version „IRT\_5keil“ S. 103; Zeilen 232 und 257*). Bei der Anpassung des Spektrums ist diese Vorgehensweise legitim, jedoch nicht bei der Bearbeitung der Koeffizienten. Hier muss auch der Gleichspannungsanteil berücksichtigt werden. Auf Grund der begrenzten Anzahl von Abtastwerten ist bei der diskreten Fourier-Transformation dem ersten DFT-Koeffizienten nicht unbedingt nur der Gleichspannungsanteil zugeordnet, sondern auch - abhängig von der Anzahl der Abtastwerte im Zeitbereich bzw. den Koeffizienten im Spektralbereich - ein Teil des niederfrequenten Bereichs.<sup>113</sup>

Beginnt die Bearbeitung der DFT-Koeffizienten in der Funktion „koef\_bearbeiten“ bereits beim ersten Koeffizienten, so entsteht beim Ausgangssignal kein Gleichspannungsanteil und die Signalsprünge im zusammengesetzten Signal können verhindert werden. Das Ergebnis zeigt Abbildung 7.2-5.



**Abbildung 7.2-5** Fenster der beiden Signale  $l[n] = 0,4 \cdot \sin(2\pi \cdot 30\text{Hz} \cdot nT)$  und  $c[n] = 0,4 \cdot \sqrt{2} \cdot \sin(2\pi \cdot 30\text{Hz} \cdot nT + \pi)$  aus dem IRT-Downmix nach Martin Keil (IRT\_5keil). Die Bearbeitung der DFT-Koeffizienten beginnt mit dem ersten Koeffizienten.

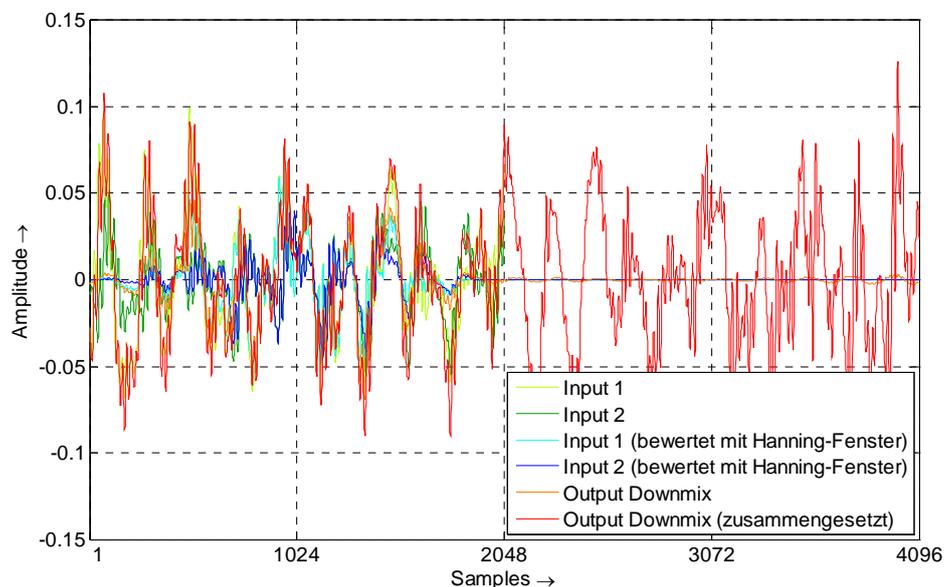
<sup>113</sup> vgl. Kapitel 4.8.1 Vom Zeit- in den Frequenzbereich

Subjektiv betrachtet bringt diese Optimierung keine wahrnehmbare Verbesserung. Die in Kapitel 7.1 beschriebenen Artefakte sind weiterhin zu hören. Dennoch ist es sinnvoll, diese Änderung zu übernehmen und den messbaren Fehler damit zu verhindern.

#### 7.2.4 Analyse mit einem stochastischen Signal

Eine zweite Analyse wird mit einem stochastischen Signal durchgeführt<sup>114</sup>. Als Testmaterial wird ein kurzer Ausschnitt aus der Matthäus-Passion (BWV 244) von Johann Sebastian Bach<sup>115</sup> verwendet.

Auf den ersten Blick erscheint es schwierig, aus dem Ergebnis (Abbildung 7.2-6) die Richtigkeit oder Unrichtigkeit des Signalverlaufs herauszulesen. Im Vergleich zu einem Sinuston wirkt die grafische Aufbereitung durch die breitbandigen und nicht vorhersehbaren Frequenzanteile des Signals abgehackt und diffus. Eventuelle Signalsprünge im zusammengesetzten Ausgangssignal sind daher nur sehr schwierig auszumachen.

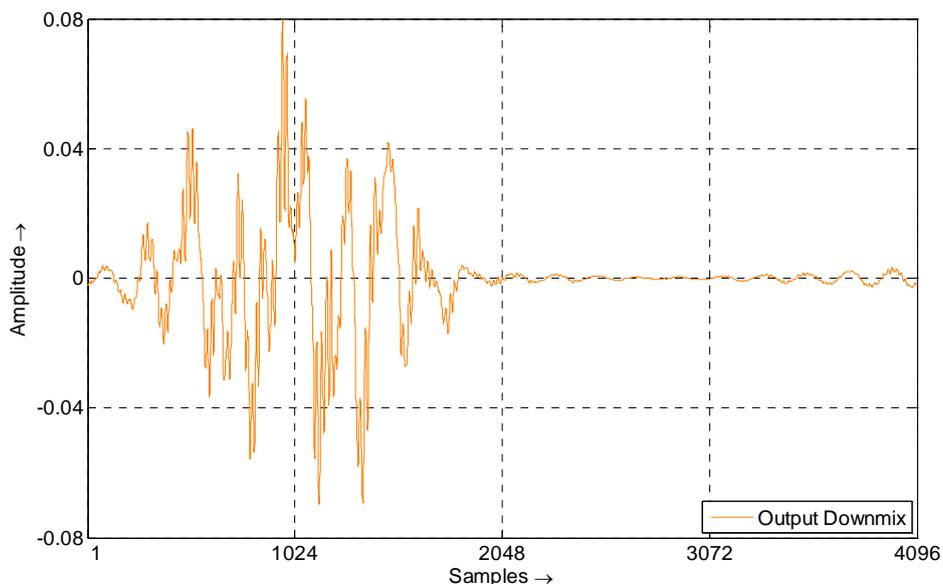


**Abbildung 7.2-6** Fenster aus dem IRT-Downmix nach Martin Keil (IRT\_5keil). Die Bearbeitung in der Funktion „koef\_bearbeiten“ beginnt mit dem ersten Koeffizienten. Das Testmaterial stammt aus der Matthäus-Passion (BWV 244) von J. S. Bach. Darstellung aller sechs Signalproben.

Betrachtet man das orangefarbene, noch gefenstertere Ausgangssignal, so fällt auf, dass trotz der Bearbeitung des ersten DFT-Koeffizienten das Signal im Bereich des Zero-Paddings nicht konstant null ist. Zur genaueren Analyse zeigt Abbildung 7.2-7 nur das Ausgangssignal.

<sup>114</sup> vgl. Grünigen 2004, S. 131

<sup>115</sup> Die Aufnahme entstand mit den „Wiener Sängerknaben“, dem „Arnold Schoenberg Chor“ und den „Concentus musicus Wien“ unter der Leitung von Nikolaus Harnoncourt. Der verwendete Ausschnitt ist aus der 2. Szene (Letztes Passahmahl wird erstes Abendmahl), in der Jesus zu den Jüngern spricht.



**Abbildung 7.2-7** Fenster aus dem IRT-Downmix nach Martin Keil (IRT\_5keil). Die Bearbeitung in der Funktion „koef\_bearbeiten“ beginnt mit dem ersten Koeffizienten. Das Testmaterial stammt aus der Matthäus-Passion (BWV 244) von J. S. Bach. Darstellung des Output-Signals.

Aus Abbildung 7.2-7 kann entnommen werden, dass es sich im zweiten Teil des Fensters offensichtlich um keinen Gleichspannungsanteil, sondern um einen ab- und anschwellenden Schwingungsvorgang handelt. Dabei verzeichnet der Bereich um das 3072. Sample den geringsten Ausschlag.

Betrachtet man den Anfang und das Ende des Fensters, so fällt auf, dass weder das erste noch das letzte Sample gleich null ist und damit, ähnlich wie bei dem Gleichspannungsanteil im vorangegangenen Kapitel, beim Zusammensetzen der Fenster ein Signalsprung verursacht wird. Bei einem stochastischen Signal ist solch ein Signalsprung grafisch nur schwer auszumachen. Er kann jedoch durchaus hörbar sein.

### 7.2.5 Ein- und Ausschwingvorgang durch die Bearbeitung

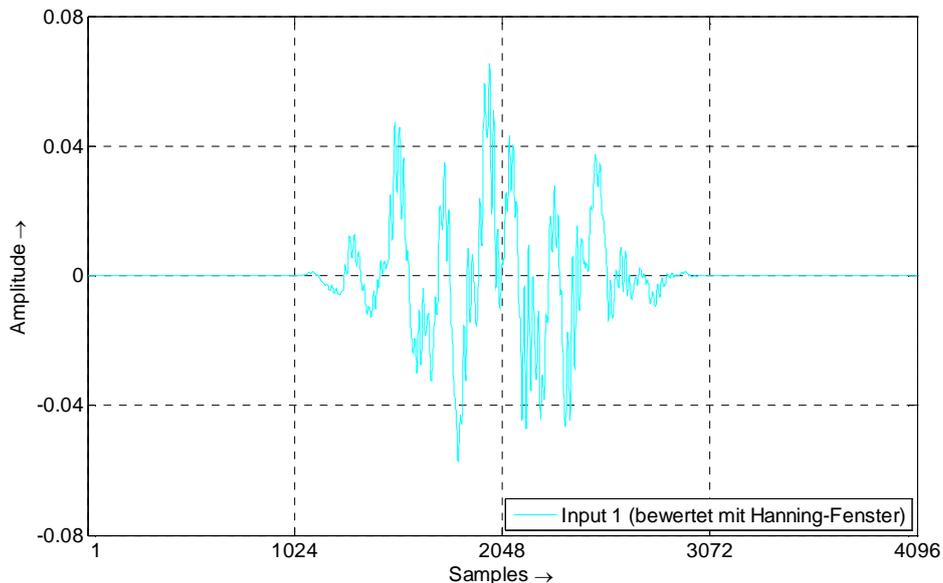
Ähnlich einem Ein- und Ausschwingvorgang entstehen die aufgespürten Schwingungen im Bereich des Zero-Paddings durch eine Faltung. Diese Faltung wird durch Multiplikationen bei der Bearbeitung der DFT-Koeffizienten im Spektralbereich hervorgerufen.<sup>116</sup> Die Multiplikation der DFT-Koeffizienten mit reellen Faktoren führt zu einer „geraden“ FIR-Filterfunktion<sup>117</sup>, die zur y-Achse symmetrisch ist. Das mit der Filterfunktion gefaltete Signalfenster wird dadurch in beide Richtungen der x-Achse sozusagen „verschmiert“. Es entsteht eine Art Ein- und Ausschwingvorgang des Filters. Ist die Hüllkurvenbewertung des Signalfensters nicht symmetrisch, liegt also das Maximum nicht in der Mitte des Fensters, so wird auf Grund der Periodizität der Fourier-Transformierten der „Einschwingvorgang“ an den „Ausschwingvorgang“ angehängt.

<sup>116</sup> vgl. Kapitel 4.8.1 Vom Zeit- in den Frequenzbereich; sowie (4.20)

<sup>117</sup> vgl. Kammeyer & Kroschel 2006, S. 158 ff.

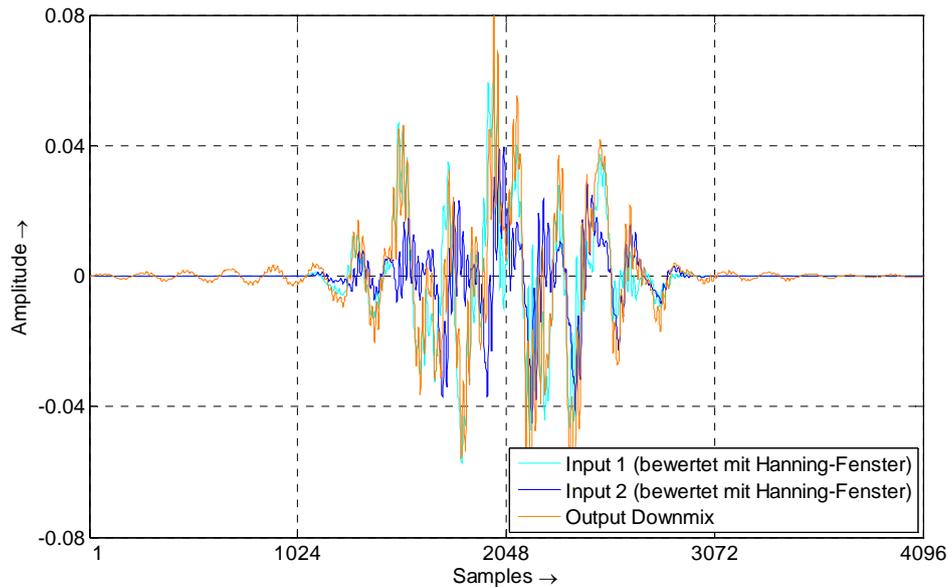
Die Schwingungen im zweiten Teil des Fensters liegen also in der durch die Bearbeitung der DFT-Koeffizienten hervorgerufenen Faltung begründet und sind demzufolge legitim. So gesehen sind sie Teil der gewünschten Korrektur.

Um die Signalsprünge beim Zusammensetzen der einzelnen Fenster zu vermeiden, muss der „Einschwingvorgang“ (3073. bis 4096. Sample in Abbildung 7.2-7) an den Anfang des Fensters gesetzt werden. Diese Änderung im Fenster ist an zwei Stellen des Additionsvorgangs möglich. Entweder beim Zusammensetzen der Fenster am Ende der Funktion „mix“ (vgl. *Quelltext: IRT-Downmix, Version „IRT\_5keil“ S. 103; Zeilen 212 und 216*) oder bereits beim Zero-Padding (*Zeilen 185-187*). Bei der zweiten Variante werden die Nullen für das Zero-Padding nicht nur an das Ende des gefensterten Eingangssignals angehängt, sondern die Hälfte der Nullen - also 1024 - auch vor dem Eingangssignal angesetzt, so dass der „Einschwingvorgang“ sofort an der richtigen Stelle erzeugt wird. Technisch gesehen liefern beide Methoden das gleiche Ergebnis. Der Übersicht wegen wird die Veränderung im Fenster schon beim Zero-Padding vorgenommen. Abbildung 7.2-8 zeigt das auf diese Weise erstellte Fenster.



**Abbildung 7.2-8** Fenster aus dem IRT-Downmix. Die Nullen des „Zero-Paddings“ werden zur Hälfte jeweils vor und nach dem gefensterten Signal angesetzt. Das Testmaterial stammt aus der Matthäus-Passion (BWV 244) von J. S. Bach. Darstellung des ersten Input-Signals.

Die grafische Auswertung des Additionsvorgangs mit der geänderten Anordnung des Zero-Paddings liefert das Diagramm in Abbildung 7.2-9. Um die Übersicht zu wahren, sind nur die mit dem Hanning-Fenster bewerteten Signalproben abgebildet.



**Abbildung 7.2-9** Fenster aus dem IRT-Downmix. Die Nullen des „Zero-Paddings“ werden zur Hälfte jeweils vor und nach dem gefensternten Signal angesetzt. Die Bearbeitung in der Funktion „koef\_bearbeiten“ beginnt mit dem ersten Koeffizienten. Das Testmaterial stammt aus der Matthäus-Passion (BWV 244) von J. S. Bach. Darstellung der beiden bewerteten Eingangssignale sowie des Ausgangssignals.

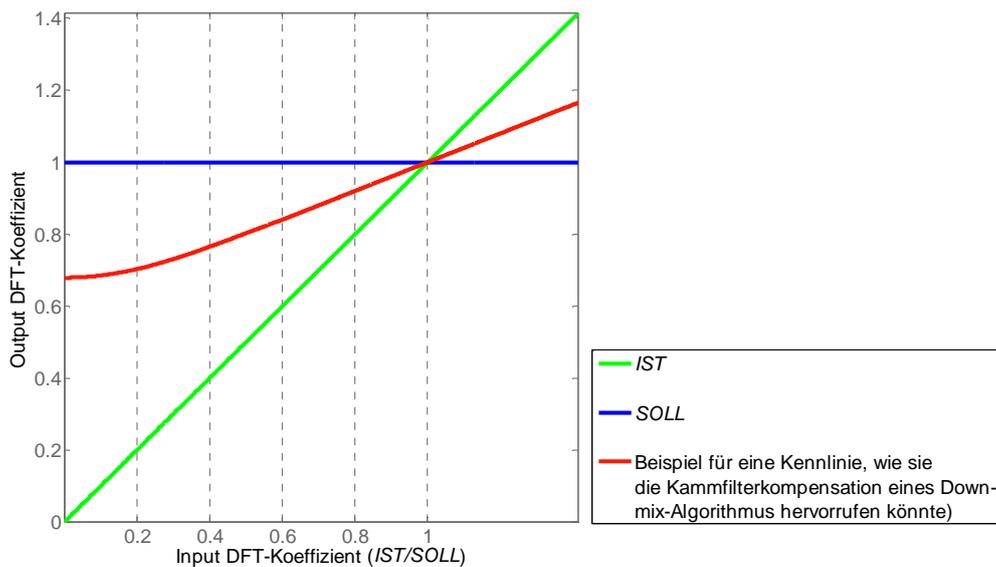
Auch subjektiv ist diese Korrektur wahrzunehmen. Die in Kapitel 7.1 beschriebenen Artefakte erfahren eine hörbare Verbesserung, wenngleich sie nach wie vor vorhanden sind. Daher ist anzunehmen, dass das fehlerhafte Zusammensetzen der bearbeiteten Fenster nicht die alleinige Ursache für die Artefakte ist.

## 7.3 Optimierung durch Analysen im Spektralbereich

Die eigentliche Logik des Downmixalgorithmus geschieht im Spektralbereich. Das Signal ist hier über DFT-Koeffizienten definiert, die mit den in Kapitel 6.1 genannten Zielen bearbeitet werden. Diese Bearbeitung im Spektralbereich soll durch eine grafische Aufbereitung objektiv greifbar und somit kontrollierbar gemacht werden.

### 7.3.1 Grafische Darstellung der Bearbeitung im Spektralbereich

Zur grafischen Beurteilung dieser Bearbeitung der DFT-Koeffizienten im Spektralbereich bietet sich ein Input/Output-Diagramm an. Ähnlich der Kennlinie bei der dynamischen Bearbeitung von Audiosignalen entsteht eine Kennlinie, die das Verstärkungsverhältnis des DFT-Koeffizienten am Ausgang in Bezug auf seinen Wert am Eingang anzeigt.



**Abbildung 7.3-1** Input/Outputdiagramm mit den Kennlinien für SOLL, IST und einem fiktiven Downmixalgorithmus mit Kammfilterkompensation

Wird das Signal nicht bearbeitet, so erhält man eine Diagonale durch den Ursprung des Diagramms, die Kennlinie eines Systems, das keine Veränderung am Signal vornimmt (siehe grüner Graph in Abbildung 7.3-1).

Eine optimale Kompensation korrigiert jeden Wert, der am Eingang anliegt, auf die Referenz, den *SOLL*-Wert. In dem Fall ist der Ausgangswert immer gleich dem *SOLL*-Wert, unabhängig davon, wie groß die Amplitude am Eingang ist. Es entsteht eine zur x-Achse parallele Gerade, welche die y-Achse auf Höhe des *SOLL*-Werts schneidet (siehe blauer Graph in Abbildung 7.3-1). Diese Gerade ist die Kennlinie einer optimalen Kammfilterkompensation. Der Schnittpunkt aus dieser Kennlinie und der Diagonalen, die entsteht, wenn das Signal nicht verändert wird, repräsentiert den Fall, bei dem *IST*- und *SOLL*-Wert gleich sind. Tritt dieser Fall ein, ist keine Bearbeitung erforderlich.

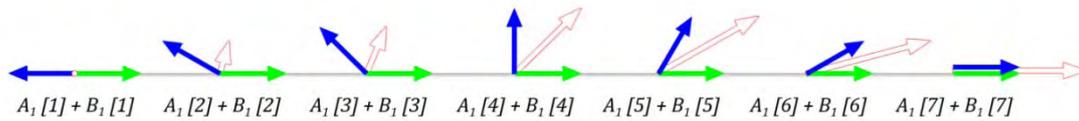
### Das Testsignal

Zur Darstellung der Kennlinie des IRT-Downmixes müssen zwei Testsignale  $A[k]$  und  $B[k]$  entworfen werden, die bei ihrer Addition alle möglichen Fälle - von der kompletten Auslöschung bis hin zur Verdopplung - berücksichtigen. Die Voraussetzung für den späteren Vergleich ist, dass unabhängig vom Ergebnis der Summe die energetische Summe, also *SOLL*, immer gleich bleibt. Ist  $SOLL = 1$ , so gilt für (6.1) und (6.2):

$$\begin{aligned} IST &= A[k] + B[k], \\ SOLL &= \sqrt{|A[k]|^2 + |B[k]|^2} \\ &= 1. \end{aligned} \tag{7.1}$$

Ein solches Testsignal lässt sich zum Beispiel relativ einfach erzeugen, wenn  $A_1[k]$  und  $B_1[k]$  zu jeder Zeit eine gleich große Amplitude aufweisen, die Phasenlage sich aber so verändert, dass

sämtliche Additionsfälle auftreten. Hinsichtlich der Phasendifferenz  $\Delta\varphi$  werden also von  $\Delta\varphi = 180^\circ$ , was eine vollständige Auslöschung hervorruft, bis hin zu  $\Delta\varphi = 0^\circ$ , was eine Verdopplung zur Folge hat, eine Reihe von Zwischenschritten generiert (Abbildung 7.3-2).



**Abbildung 7.3-2** Testsignalreihen  $A_1[k]$  (blau) und  $B_1[k]$  (grün), mit deren Hilfe die Kennlinie des IRT-Downmix berechnet wird. Der weiße Pfeil mit roter Kontur stellt die Summe der beiden Signale dar. Die Entwicklung der Reihen basiert auf festen Amplituden  $|A_1[k]| = |B_1[k]|$  und einem variablen Phasenwinkel  $0^\circ \leq \angle A_1[k] \leq 180^\circ$ .

Sind beide Amplituden der zu summierenden Signale gleich groß  $|A_1[k]| = |B_1[k]|$ , und soll ihre energetische Summe 1 betragen, so ergibt (7.1) nach Umformung für die Amplituden den Wert:

$$\begin{aligned} |A_1[k]| = |B_1[k]| &= \frac{1}{\sqrt{2}} \\ &= 0,7071. \end{aligned}$$

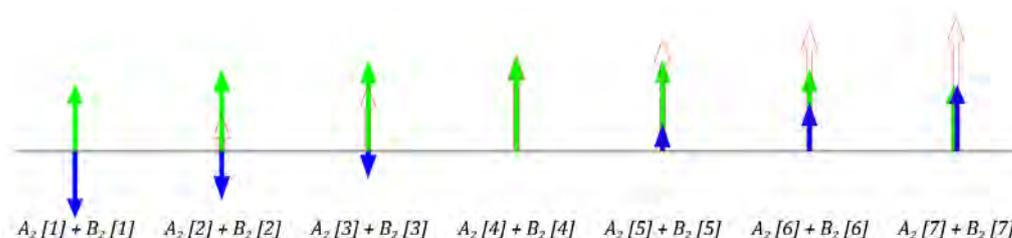
Die Testsignalkombination besteht also aus einer konstanten Amplitude mit dem Wert 0,7071,  $B_1[k]$  hat den konstanten Phasenwinkel  $\angle B_1[k] = 0^\circ$  und  $A_1[k]$  einen variablen Phasenwinkel im Bereich von  $0^\circ \leq \angle A_1[k] \leq 180^\circ$ .

Eine zweite Möglichkeit, eine Testsignalkombination  $A_2[k]$  und  $B_2[k]$  zu erstellen, besteht darin die Amplituden zu variieren und die Phasenwinkel konstant zu halten:  $\angle A_2[k] = \angle B_2[k] = 0^\circ$ . Ist beispielsweise das eine Signal  $A_2[k] = 0$ , so muss das andere  $B_2[k] = SOLL = 1$  sein, so dass die Bedingung (7.1) erfüllt ist.

Generell lässt sich aus (7.1) die Amplitude  $|B_2[k]|$  in Abhängigkeit von  $|A_2[k]|$  errechnen:

$$|B_2[k]| = \sqrt{SOLL - |A_2[k]|^2} \quad (7.2)$$

Um alle Additionsfälle abzudecken, wird  $|A_2[k]|$  von -0,7071 bis 0,7071 schrittweise erhöht. Die zugehörigen Werte für  $|B_2[k]|$  werden über (7.2) berechnet. Abbildung 7.3-3 veranschaulicht die auf diese Weise erstellten Testreihen  $A_2[k]$  und  $B_2[k]$ .



**Abbildung 7.3-3** Testsignalreihen  $A_2[k]$  (blau) und  $B_2[k]$  (grün), mit deren Hilfe die Kennlinie des IRT-Downmix berechnet wird. Der weiße Pfeil mit roter Kontur stellt die Summe der beiden Signale dar. Die Entwicklung der Reihen basiert auf variablen Amplituden und einer konstanten Phasenlage.

Beide Testsignalkombinationen stellen Extremsituationen dar. Die erste provoziert alle Additionsfälle über einen variablen Phasenwinkel und eine feste Amplituden, die zweite über variable Amplituden und einen festen Phasenwinkel. In der Praxis auftretende Signalkombinationen sind entweder von der Art der ersten oder der zweiten Testsignalkombination oder in den allermeisten Fällen eine Mischung aus beiden. Aus diesem Grund werden für die Analyse beide Signalkombinationen verwendet.

Zur Bestätigung dieser theoretischen Überlegungen, wird in dem Diagramm ebenfalls die Bearbeitung eines stochastischen Signals ausgewertet. Die daraus resultierende „reale Verteilung“ sollte theoretisch von den beiden Kennlinien der Testsignale eingeschlossen werden.

### Realisierung der grafischen Darstellung

Das Programm „*dft\_koef\_darstellung.m*“<sup>118</sup> generiert die beiden Testsignalkombinationen, führt mit ihnen und dem stochastischen Signal einen Additionsvorgang des Downmixverfahrens durch und stellt das Ergebnis beider Signalkombinationen als Kennlinie dar, das des stochastischen Signals als reale Verteilung. Zur Orientierung werden auch die Linien des *SOLL*- und *IST*-Zustands eingezeichnet. Abbildung 7.3-4 zeigt das Diagramm für den IRT-Downmix auf dem Stand von Martin Keil.

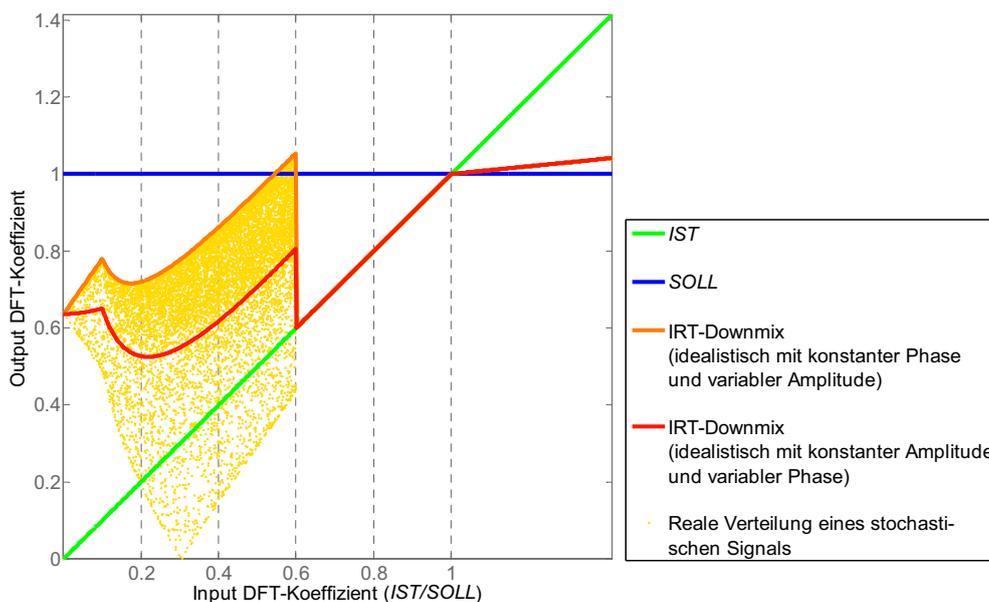


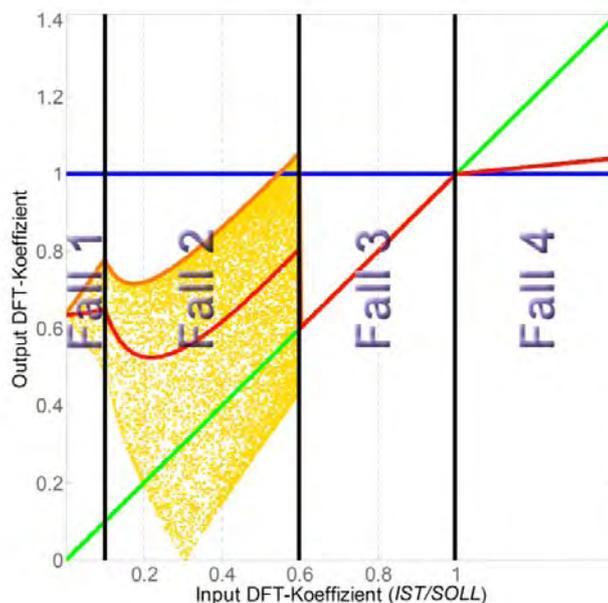
Abbildung 7.3-4 Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix nach Martin Keil „IRT\_5keil“

### 7.3.2 Analyse der Bearbeitung der DFT-Koeffizienten

Die zwei Kennlinien für die Bearbeitung der DFT-Koeffizienten des IRT-Downmixes nach Martin Keil ist in Abbildung 7.3-1 ersichtlich. Die Kennlinie wirkt auf den ersten Blick eckig und abgehackt. Der Verlauf bei den Werten 0,1; 0,6 und 1 für *IST/SOLL* lässt klar die drei Fallunterscheidungen

<sup>118</sup> Der Quelltext des Programms „*dft\_koef\_darstellung.m*“ ist im Anhang auf S. 121 zu finden.

erkennen. Auch die vier unterschiedlichen Fälle der Bearbeitung sind deutlich auszumachen (Abbildung 7.3-5).<sup>119</sup>



**Abbildung 7.3-5** Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix nach Martin Keil. Beschriftung der Fälle unterschiedlicher Bearbeitung.

Im Bereich von Fall 4 wird wie erwartet der zu starke Anstieg auf 10% reduziert. Liegt der Wert am Eingang zwischen 0,6 und 1, so tritt Fall 3 und somit keine Bearbeitung der DFT-Koeffizienten in Kraft. Auch dieser Bereich liefert das erwünschte Ergebnis, im Gegensatz zu den Fällen 1 und 2. Die lineare Verstärkung des ersten Eingangssignals in Fall 2 führt zu einem nach oben geöffneten, parabelförmigen Verlauf, so dass Eingangswerte mit einem IST/SOLL-Verhältnis von 0,1 eine größere Verstärkung erfahren, als beispielsweise Werte mit einem Verhältnis von 0,2. Das bedeutet, die Verstärkung und damit die Kompensation in diesem Abschnitt verlaufen nicht proportional zum Grad der Auslöschung.

Der Übergang von Fall 2 zu Fall 3 ist nicht stetig, so dass ein Sprung in der Kennlinie entsteht. Es wäre nicht verwunderlich, wenn ein DFT-Koeffizient, welcher an dieser Stelle im Diagramm über mehrere Fenster alternierend von Fall 2 zu Fall 3 hin und her springt, im zusammengesetzten Ausgangssignal Artefakte auslösen würde. Nicht auszuschließen ist dieser Effekt auch beim Übergang von Fall 1 zu Fall 2, bei dem die Kennlinie zwar keinen Sprung aufweist, jedoch ein Knick.

Betrachtet man die reale Verteilung des stochastischen Signals, so fällt auf, dass die Testsignalkombination mit konstanter Phase und variabler Amplitude als Extremsituation wie erwartet die Werte von oben her einschließt. Die Signalkombination mit fester Amplitude und variabler Phasendifferenz hingegen schließt die reale Verteilung nach unten hin nicht ab. Auch wenn eine Häufung der Werte des stochastischen Signals zwischen den beiden Kennlinien der Testsignalkombinationen auszumachen ist, entsteht offensichtlich bei einigen Werten sogar eine Dämpfung. Es handelt sich dabei um die Punkte, die unter der *IST*-Linie liegen, also schlechter abschneiden, als wenn sie nicht bearbeitet worden wären. Auf die Kompensation von Kammfiltern projiziert bedeutet dies, dass in bestimmten Fällen der Kammfiltereffekt nicht korrigiert, sondern verstärkt wird.

<sup>119</sup> vgl. Kapitel 6.2.2 Der Algorithmus

Die erste Auswertung der Kennlinien des IRT-Downmixes und der realen Verteilung, die aus der Bearbeitung eines stochastischen Signals resultiert, wirft einige Fragen auf und deutet auf eventuelle Fehler hin, die zu überprüfen und gegebenenfalls zu korrigieren sind.

Als erstes soll versucht werden, durch eine Optimierung der Kennlinie mögliche, durch den unsteuerten Verlauf der Kennlinie verursachte Artefakte zu eliminieren.

### 7.3.3 Optimierung der Kennlinie

Die problematischen Übergänge zwischen den einzelnen Bearbeitungsfällen der DFT-Koeffizienten legen nahe, die Anzahl der Bearbeitungsmöglichkeiten zu verkleinern, also die vier Fälle teilweise - je nach Möglichkeit - zusammenzufassen. Durch eine geschickte Veränderung der Bearbeitung in Fall 2 wäre es durchaus denkbar Fall 1 in diesen zu integrieren. Auch Fall 3, bei dem in Keils Version die DFT-Koeffizienten nicht verändert werden und dadurch beim Übergang zu Fall 2 einen kritischen Sprung in der Kennlinie verursachen, könnte entweder in den Fall 2 oder Fall 4 übernommen werden. Dies würde zwei der vier Fälle im Algorithmus überflüssig machen und hätte nur noch eine Fallunterscheidung bzw. einen Übergang in der Kennlinie zur Folge.

Für eine solche Veränderung ist eine Anpassung der Bearbeitung in den Fällen 2 und 4 allerdings unausweichlich. Besonders Fall 2 stellt eine Herausforderung dar, da die lineare Bearbeitung der Amplitude des ersten DFT-Koeffizienten, wie sie in Fall 2 vorgenommen wird, keine lineare Veränderung der Summe der beiden Eingangssignale zur Folge hat. Der Zusammenhang des Korrekturfaktors und der Veränderung in der resultierenden Summe ist theoretisch herzuleiten.

#### Theoretische Überlegungen zur Optimierung der Kennlinie

Theoretisch ist es möglich, eine Absenkung der DFT-Koeffizienten der Summe, die bisher von den Fällen 1-3 bearbeitet werden, vollständig auszugleichen. Die Beziehung der beiden Eingangssignale  $A[k]$  und  $B[k]$ , ihrer korrigierten Summe  $S_{\text{korrigiert}}[k]$  und dem Referenzwert  $SOLL$  definiert (7.3).

$$\begin{aligned} S_{\text{korrigiert}}[k] &= x \cdot A[k] + B[k], \\ \text{wobei } |S_{\text{korrigiert}}[k]| &= SOLL \\ \text{und } x &\in \mathbb{R}. \end{aligned} \quad (7.3)$$

Die Größe des reellen Korrekturfaktors  $x$  ist dabei nicht nur von der Größe der Amplituden der beiden Eingangssignale, sondern auch von deren Phasenwinkel abhängig. Soll  $x$  in Abhängigkeit der beiden DFT-Koeffizienten  $A[k]$  und  $B[k]$  am Eingang bestimmt werden, so löst man (7.3) nach  $SOLL$  auf:

$$SOLL = |x \cdot A[k] + B[k]|. \quad (7.4)$$

Mit der Definition für  $SOLL$  aus (6.2) erhält man:

$$\sqrt{|A[k]|^2 + |B[k]|^2} = |x \cdot A[k] + B[k]|. \quad (7.5)$$

Auf der rechten Seite steht der Betrag einer komplexen Zahl  $x \cdot A[k] + B[k]$ . Anhand der Definition (4.12) besteht der Betrag einer komplexen Zahl aus der Hypotenuse zu Real- und Imaginärteil der komplexen Zahl.

$$\sqrt{|A[k]|^2 + |B[k]|^2} = \sqrt{\operatorname{Re}(x \cdot A[k] + B[k])^2 + \operatorname{Im}(x \cdot A[k] + B[k])^2}. \quad (7.6)$$

Links unter der Wurzel kann  $|B[k]|^2$  nach Definition (4.12) als  $\operatorname{Re}(B[k])^2 + \operatorname{Im}(B[k])^2$  geschrieben werden. Ebenso wird  $\operatorname{Re}(x \cdot A[k] + B[k]) = x \cdot \operatorname{Re}(A[k]) + \operatorname{Re}(B[k])$  und  $\operatorname{Im}(x \cdot A[k] + B[k]) = x \cdot \operatorname{Im}(A[k]) + \operatorname{Im}(B[k])$  umgeformt.

Quadrieren der Gleichung und Ausmultiplizieren ergibt:

$$\begin{aligned} & |A[k]|^2 + \operatorname{Re}(B[k])^2 + \operatorname{Im}(B[k])^2 \\ &= (x \operatorname{Re}(A[k]))^2 + 2x \operatorname{Re}(A[k]) \operatorname{Re}(B[k]) + (\operatorname{Re}(B[k]))^2 \\ &+ (x \operatorname{Im}(A[k]))^2 + 2x \operatorname{Im}(A[k]) \operatorname{Im}(B[k]) + (\operatorname{Im}(B[k]))^2. \end{aligned} \quad (7.7)$$

Vereinfachen und Sortieren führt zu einer quadratischen Gleichung:

$$(\operatorname{Re}(A[k])^2 + \operatorname{Im}(A[k])^2)x^2 + 2(\operatorname{Re}(A[k])\operatorname{Re}(B[k]) + \operatorname{Im}(A[k])\operatorname{Im}(B[k]))x - |A[k]|^2 = 0. \quad (7.8)$$

Zur besseren Übersicht werden beide Seiten durch  $\operatorname{Re}(A[k])^2 + \operatorname{Im}(A[k])^2 = |A[k]|^2$  geteilt:

$$x^2 + 2 \left( \frac{\operatorname{Re}(A[k])\operatorname{Re}(B[k]) + \operatorname{Im}(A[k])\operatorname{Im}(B[k])}{|A[k]|^2} \right) x - 1 = 0, \quad (7.9)$$

und zur Lösung die Mitternachtsformel<sup>120</sup> angewandt:

$$\begin{aligned} x_{1/2} &= \frac{-2p \pm \sqrt{4p^2 + 4}}{2}, \\ \text{wobei } p &= \left( \frac{\operatorname{Re}(A[k])\operatorname{Re}(B[k]) + \operatorname{Im}(A[k])\operatorname{Im}(B[k])}{|A[k]|^2} \right). \end{aligned} \quad (7.10)$$

Schließlich kann die 2 im Nenner gekürzt werden und man erhält die überschaubare Lösung:

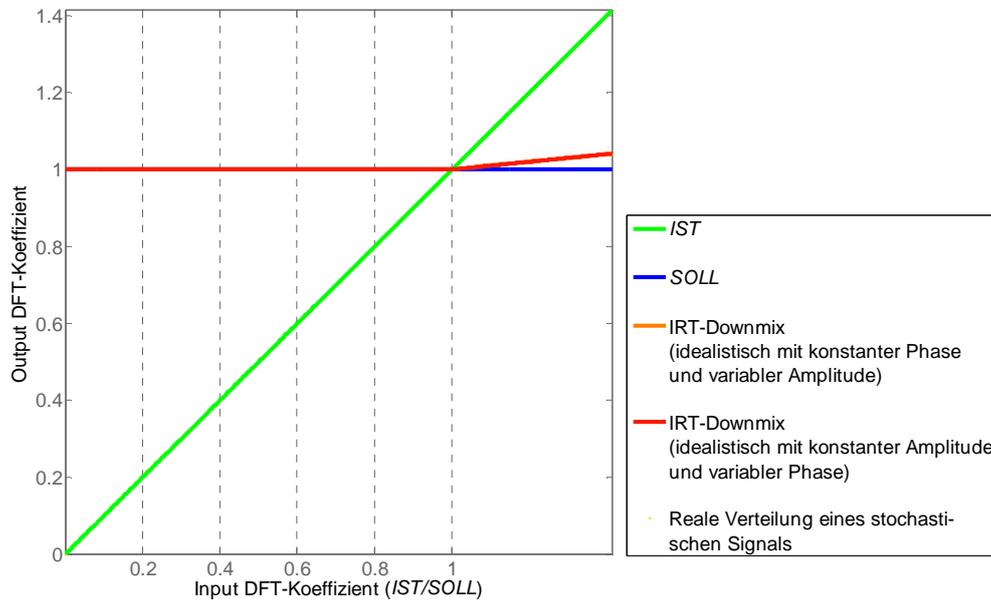
$$\begin{aligned} x_{1/2} &= -p \pm \sqrt{p^2 + 1}, \\ \text{wobei } p &= \frac{\operatorname{Re}(A[k])\operatorname{Re}(B[k]) + \operatorname{Im}(A[k])\operatorname{Im}(B[k])}{|A[k]|^2} \end{aligned} \quad (7.11)$$

Wird der erste der beiden DFT-Koeffizienten am Eingang  $A[k]$  mit dem Korrekturfaktor  $x$  bearbeitet, so kann eine Auslöschung vollständig kompensiert werden.

Die Bearbeitung mit dem Korrekturfaktor aus (7.11) kann für sämtliche DFT-Koeffizienten, die in Keils Algorithmus den Fällen 1-3 zugeordnet sind, angewandt werden. Für DFT-Koeffizienten, deren IST-Wert kleiner als der SOLL-Wert ist, gilt der neue Fall 1 ( $IST/SOLL < 1$ ). Für alle anderen ( $IST/SOLL \geq 1$ ) Fall 2, der identisch mit Fall 4 aus Keils Version ist.

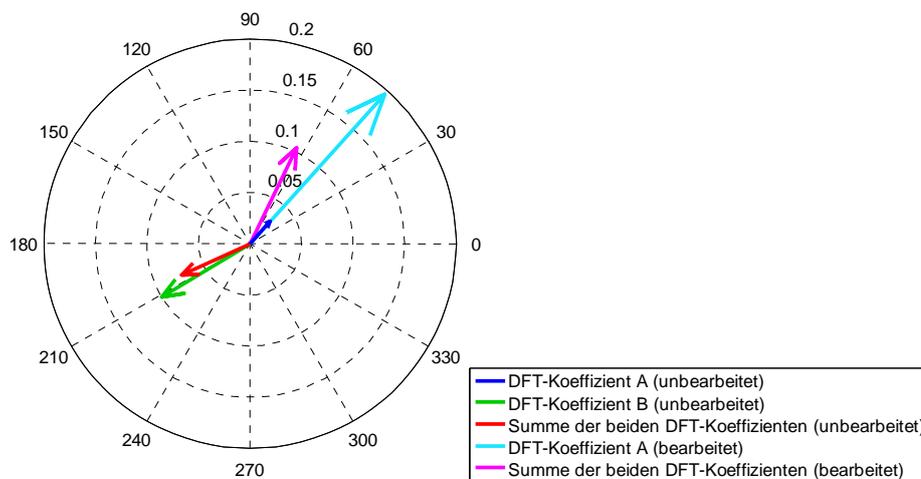
<sup>120</sup>  $x_{1/2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ , bei  $ax^2 + bx + c = 0$ .

Auch wenn die Kennlinie das gewünschte Ergebnis liefert, eine zu jedem Zeitpunkt optimale Korrektur vorweist und die Realverteilung auch tatsächlich der Kennlinie folgt (Abbildung 7.3-6), so tritt bei Verwendung dieses Korrekturfaktors eine extreme Artefaktbildung ein, die als unüberhörbares, andauerndes Knacken wahrzunehmen ist und den Ton regelrecht zerstört.



**Abbildung 7.3-6** Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix. Die Bearbeitung erfolgt mit einer Fallunterscheidung. Die Bearbeitung in Fall 1 ( $IST < SOLL$ ) erfolgt mit dem Korrekturfaktor  $x[p]$  aus (7.11).

Die Ursache für dieses unbrauchbare Ergebnis liegt in dem extremen Korrekturfaktor, der von DFT-Koeffizient zu DFT-Koeffizient zudem stark variieren kann. Zwar wird dadurch die Amplitude auf den gewünschten Wert angehoben, der Phasenwinkel erfährt im gleichen Zug jedoch für jedes Fenster eine neue, dem Signal nicht angepasste und dadurch fehlerbehaftete Änderung.



**Abbildung 7.3-7** Darstellung der an einem Additionsvorgang des IRT-Downmixes beteiligten DFT-Koeffizienten in der komplexen Ebene. Die Bearbeitung des DFT-Koeffizienten A erfolgt mit dem Korrekturfaktor  $p = 6,75$ , der mit Hilfe von (7.11) errechnet wird. Der Phasenwinkel wird radial, die Amplitude zentrifugal dargestellt.  $A[r] = 0,029 \cdot e^{0,84 \cdot i}$  und  $B[r] = 0,1 \cdot e^{3,69 \cdot i}$ .

Die Abbildung 7.3-7 zeigt, unter welchen Umständen solch ein extrem großer Korrekturfaktor entsteht: Nämlich dann, wenn der erste, zu bearbeitende DFT-Koeffizient der beiden Eingangssignale eine deutlich kleinere Amplitude aufweist als der zweite. Der in einem solchen Fall notwendige Korrekturfaktor muss die Amplitude des ersten DFT-Koeffizienten auf über die doppelte Größe des zweiten Koeffizienten verstärken. Der Winkel der resultierenden Summe ist sowohl vom Phasenwinkel als auch von der Amplitudengröße der beiden zu summierenden DFT-Koeffizienten abhängig. Die Verlängerung der ersten Amplitude bewirkt somit auch eine Änderung des Phasenwinkels der Summe.

### Der Algorithmus mit einer Fallunterscheidung

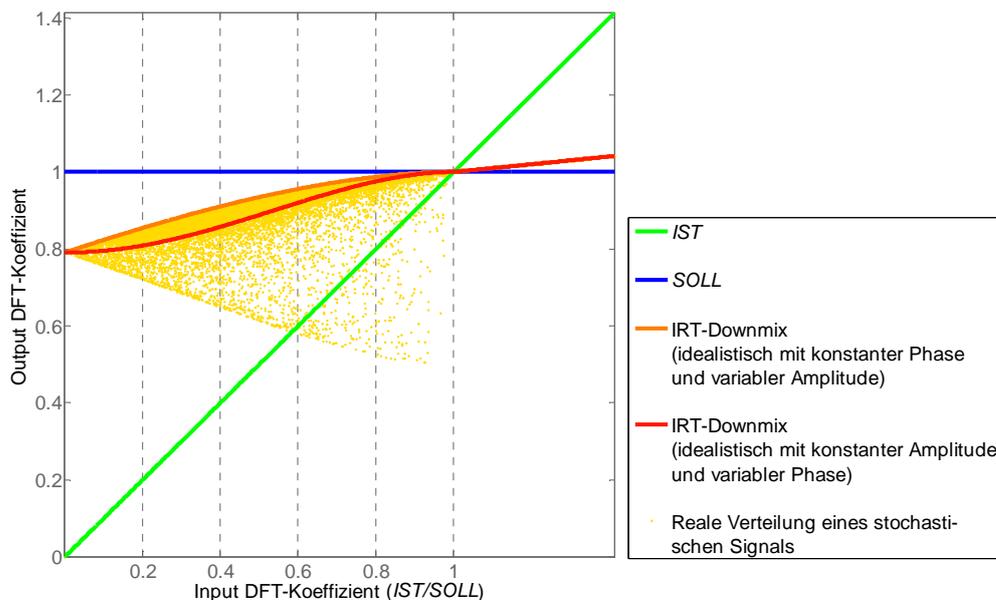
Der Grund für den extremen Korrekturfaktor  $x$  in (7.11) ist der DFT-Koeffizient  $A[k]$  im Nenner bei der Berechnung von  $p$ . Je kleiner  $A[k]$  im Nenner von  $p$  ist, desto größer wird der Betrag von  $x$ . Durch Umformung von (7.11)<sup>121</sup>, kann  $A[k]$  direkt in die nach  $x$  aufgelöste Gleichung gezogen werden, wo durch eine Gewichtung von  $p^2$  das Problem entschärft werden kann. Aus diesen Überlegungen folgt für  $x$  und den Skalierungsfaktor  $w$ :

$$x_{1/2} = \frac{-p \pm \sqrt{w \cdot p^2 + |A[k]|^4}}{|A[k]|^2}, \quad (7.12)$$

wobei  $p = \text{Re}(A[k])\text{Re}(B[k]) + \text{Im}(A[k])\text{Im}(B[k])$ .

Für den Skalierungsfaktor  $w$  hat sich ein Wert von 0,25 als sinnvoll erwiesen. Damit wird zwar keine vollständige Korrektur der DFT-Koeffizienten der Summe mehr erreicht, dafür sollte die Artefaktbildung aber erheblich geringer ausfallen.

In Fall 1 des modifizierten Algorithmus wird der Korrekturfaktor nun mit Hilfe von (7.12) ermittelt. Fall 2 bleibt unverändert. Die grafische Auswertung zeigt Abbildung 7.3-8.



**Abbildung 7.3-8** Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix „IRT\_5i“. Die Bearbeitung erfolgt mit einer Fallunterscheidung. Die Korrektur in Fall 1 ( $IST < SOLL$ ) erfolgt mit dem Korrekturfaktor  $x[p]$  aus (7.12), wobei  $w = 0,25$ .

<sup>121</sup> Diese Umformung wird am deutlichsten, löst man (7.8) direkt nach  $x$  auf.

Wie erwartet zeigen die beiden Kennlinien keine vollständige Korrektur der Absenkung im linken Teil des Diagramms. Eine Korrektur auf 80% des *SOLL*-Werts bei einer vollständigen Auslöschung ( $IST/SOLL = 0$ ) ist dennoch ein völlig ausreichendes Ergebnis. Der Verlauf der Kennlinie enthält weder Sprünge noch Ecken. Auch der Übergang von Fall 1 zu Fall 2 ( $IST/SOLL = 1$ ) ist stetig. Die Ergebniswerte des stochastischen Signals liefern zwar optisch gesehen eine schönere Verteilung als bei Keils Version, dies darf allerdings nicht darüber hinweg täuschen, dass im mittleren Teil des Diagramms ( $\sim 0,6 < IST/SOLL < 1$ ) einige bearbeitete DFT-Koeffizienten gegenüber dem *IST*-Wert auch hier eine Dämpfung erfahren.

Der subjektive Eindruck beim Abhören des auf diese Weise herunter gemischten Tonmaterials ist gut. Die Artefakte sind nicht mehr wahrnehmbar. Die Klangfarbe ist ausgeglichen. Diese Variante des IRT-Downmix erhält den Namen „IRT\_5i“.

### 7.3.4 Analyse der Dämpfung von DFT-Koeffizienten beim Bearbeitungsvorgang

Mit der Glättung der Kennlinie ist zwar eine subjektiv wahrnehmbare Verbesserung bezüglich der Artefakte des Downmixalgorithmus einhergegangen, die Dämpfung einiger DFT-Koeffizienten im Bereich  $\sim 0,6 < IST/SOLL < 1$  konnte jedoch nicht verhindert werden (Abbildung 7.3-8 im Vergleich zu Abbildung 7.3-4). Diese unerwartete negative Verstärkung lässt sich mit dem Vorsatz des IRT-Downmixes, Kammfilter bestmöglich zu korrigieren, nicht vereinbaren.

Um die Ursache für diesen unschönen Effekt herauszufinden, sind die speziellen Fälle der beiden zu summierenden DFT-Koeffizienten  $A[k]$  und  $B[k]$ , die nach dem Bearbeitungsvorgang eine solche Dämpfung aufweisen, detailliert zu betrachten.

#### Fallanalyse

Aus einem in den Spektralbereich transformierten Fenster des stochastischen Beispielsignals wird ein zu summierendes DFT-Koeffizienten-Paar  $A[k = r]$  und  $B[k = r]$  ausgesucht, deren Summe  $S_{bearbeitet}[k = r]$  nach dem Additions- und Bearbeitungsvorgang eine kleinere Amplitude aufweist als der resultierende DFT-Koeffizient einer Addition ohne Bearbeitung  $S[k = r] = IST[k = r]$ .  $S_{bearbeitet}[k = r]$  muss in Abbildung 7.3-8 also unter der grünen Kennlinie der *IST*-Werte liegen.

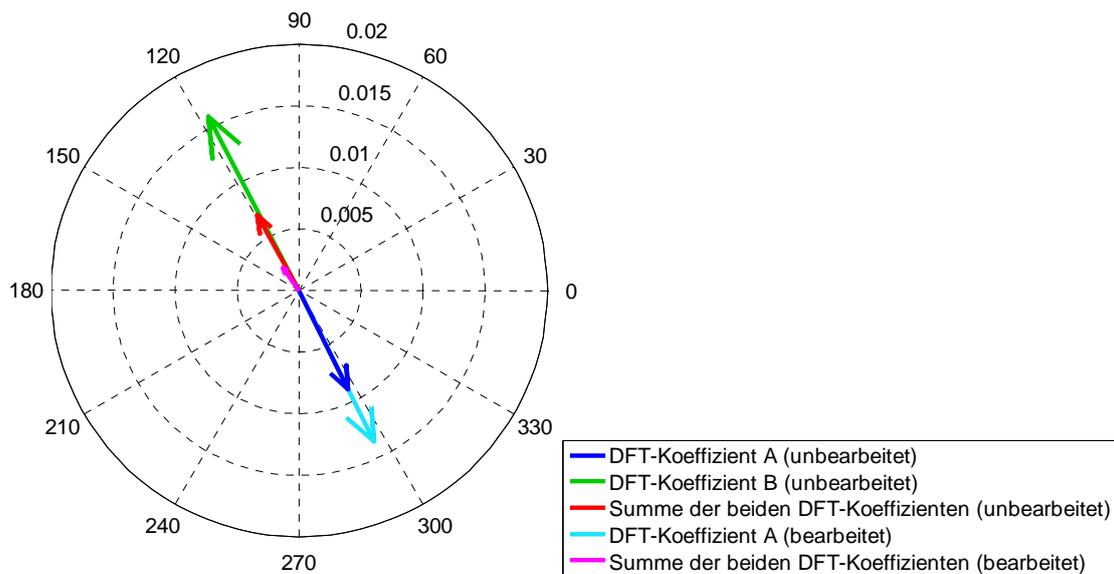
Das konkrete und im Folgenden zu analysierende Beispiel besteht aus:

$$A[r] = 0,0089 \cdot e^{5,17 \cdot i} \text{ und} \\ B[r] = 0,0159 \cdot e^{2,05 \cdot i},$$

so dass sich für die Amplituden und Winkel folgende Werte ergeben:

$$|A[r]| = 0,0089 \text{ und} \\ |B[r]| = 0,0159, \text{ sowie} \\ \angle A[r] = \frac{5,17}{\pi} \cdot 180^\circ = 296,2^\circ \text{ und} \\ \angle B[r] = \frac{2,05}{\pi} \cdot 180^\circ = 117,5^\circ.$$

Abbildung 7.3-9 zeigt die beiden zu summierenden Koeffizienten, deren Summe sowie den bearbeiteten Koeffizienten  $A_{bearbeitet}[r]$  und die korrigierte Summe in der komplexen Ebene.



**Abbildung 7.3-9** Darstellung der an einem Additionsvorgang des IRT-Downmixes nach Martin Keil beteiligten DFT-Koeffizienten in der komplexen Ebene. Der Phasenwinkel wird radial, die Amplitude zentrifugal dargestellt.  $A[r] = 0,0089 \cdot e^{5,17 \cdot i}$  und  $B[r] = 0,0159 \cdot e^{2,05 \cdot i}$ .

Die bildliche Darstellung veranschaulicht das Problem. Die Phasendifferenz der beiden zu summierenden Koeffizienten beträgt ungefähr  $180^\circ$ . Da die Amplitude von  $A[r]$  kleiner ist als die von  $B[r]$ , tritt bei einer normalen Addition eine Absenkung auf. Wird zur Kompensation dieser Absenkung der erste Koeffizient  $A[r]$  verstärkt, vergrößert sich die Absenkung, bis sie bei  $|A_{bearbeitet}[r]| = |B[r]|$  in einer vollständigen Auslöschung gipfelt. Erst wenn die Verstärkung eine doppelt so große Amplitude des ersten Koeffizienten gegenüber dem zweiten zur Folge hat, kann von einer Kompensation gesprochen werden. Eine solche Verstärkung, die mit dem Korrekturfaktor aus (7.11) realisiert wird, hat allerdings eine ungünstige Auswirkung auf den Phasenwinkel der resultierenden Summe  $S_{bearbeitet}[r]$ .<sup>122</sup> Störende Artefakte sind die Folge.

An dieser Stelle wird auch der Zusammenhang zwischen der optimalen Kompensation durch (7.11) und den Bearbeitungsmethoden deutlich, die eine Dämpfung der DFT-Koeffizienten hervorrufen. Die Divergenz und die daraus resultierende Dämpfung der realen Verteilung entstehen, wenn der Betrag des zu bearbeitenden Koeffizienten  $|A[k]|$  kleiner ist als  $|B[k]|$  und deren Phasendifferenz zwischen  $90^\circ$  und  $180^\circ$  liegt. Je stärker sich die zu summierenden DFT-Koeffizienten einer Phasendifferenz von  $180^\circ$  annähern, desto unzureichender wird die Verstärkung des kleineren, ersten Koeffizienten (Abbildung 7.3-9) und desto größer die Dämpfung gegenüber einer Addition ohne Bearbeitung.

Nur wenn der erste Koeffizient  $A[k]$  auf eine vollständige Kompensation hin korrigiert wird, kann auch die Divergenz der realen Verteilung verhindert werden. Dies ist der Grund, warum nur bei der Bearbeitung mit dem in (7.11) errechneten Kompensationsfaktor beide Kennlinien übereinander liegen und die reale Verteilung genau deren Verlauf ohne Abweichung folgt und damit

<sup>122</sup> Vgl. Kapitel 7.3.3 Theoretische Überlegungen zur Optimierung der Kennlinie

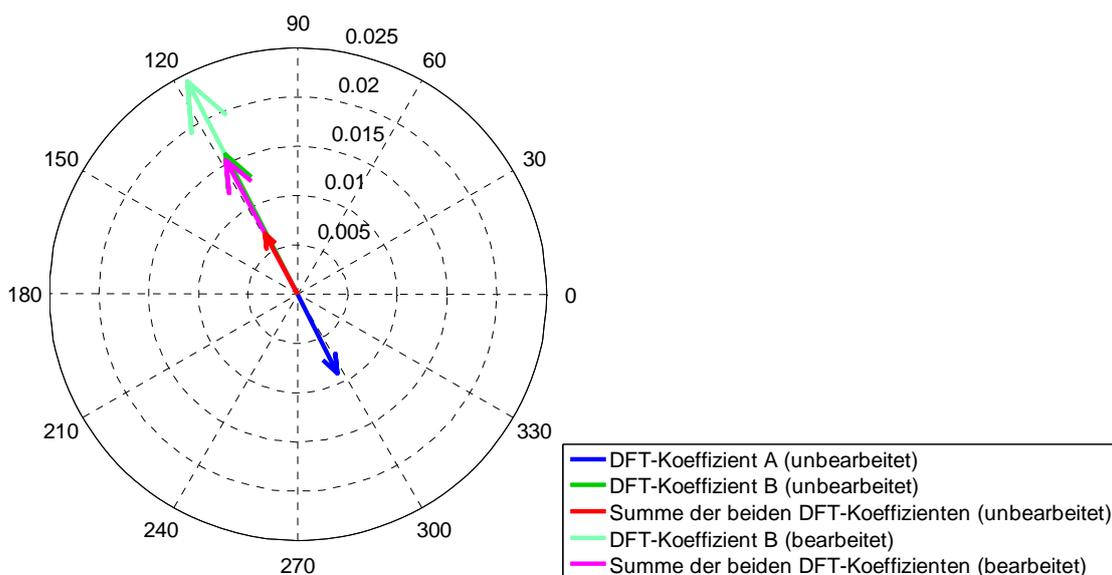
nicht divergiert. Die Werte des stochastischen Signals gelangen so nicht in den Dämpfungsbereich unter der grünen *IST*-Linie (Abbildung 7.3-6).

Das nächste Kapitel erörtert die Möglichkeit, nicht generell den ersten DFT-Koeffizienten zu bearbeiten, sondern durch eine optionale Bearbeitung des zweiten Koeffizienten an Stelle des ersten den kritischen Fällen auszuweichen.

### Bearbeitung des größeren DFT-Koeffizienten

Die optimale Korrektur des ersten DFT-Koeffizienten mit Hilfe des Korrekturfaktors aus (7.11) ist nicht die einzige Möglichkeit, das Problem der Dämpfung zu vermeiden. Die Analyse von Abbildung 7.3-9 legt nahe, nicht generell den ersten der beiden zu summierenden DFT-Koeffizienten zu bearbeiten, sondern immer denjenigen mit der größeren Amplitude. Damit würden weder eine Dämpfung noch extrem große Korrekturfaktoren entstehen.

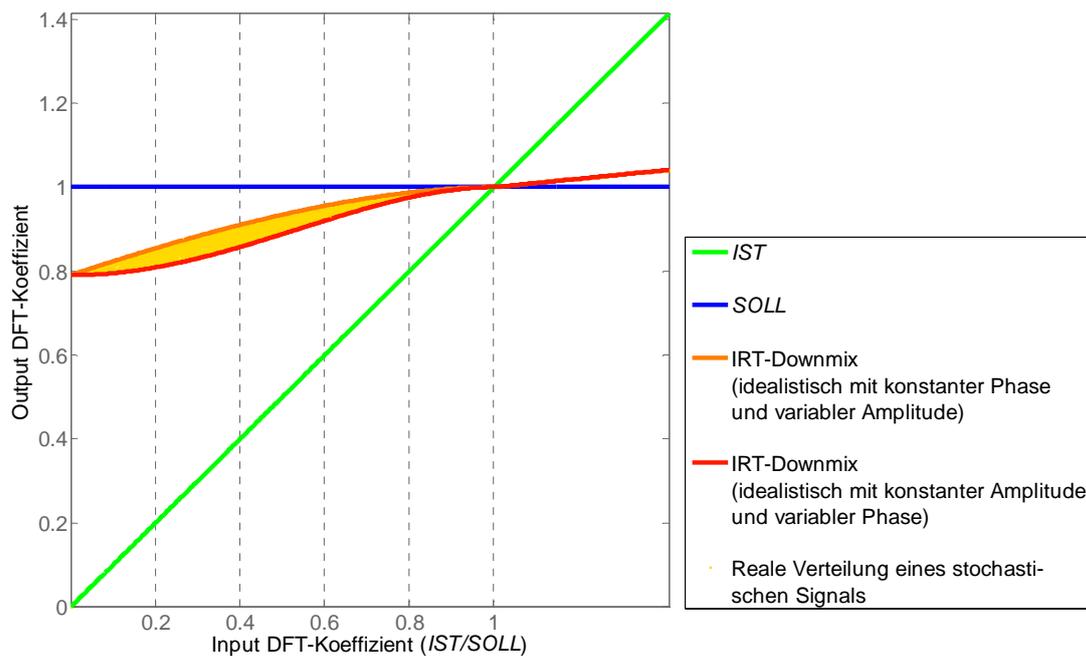
Abbildung 7.3-10 zeigt die Addition der auf diese Weise bearbeiteten DFT-Koeffizienten aus dem Beispiel im vorangegangenen Kapitel.



**Abbildung 7.3-10** Darstellung der an einem Additionsvorgang des IRT-Downmixes beteiligten DFT-Koeffizienten in der komplexen Ebene. Der Algorithmus bearbeitet den DFT-Koeffizienten mit der größeren Amplitude. Der Phasenwinkel wird radial, die Amplitude zentrifugal dargestellt.  $A[r] = 0,0089 \cdot e^{5,17 \cdot i}$  und  $B[r] = 0,0159 \cdot e^{2,05 \cdot i}$ .

Das Diagramm mit den Kennlinien und der Realverteilung (Abbildung 7.3-11) zeigt ein vielversprechendes Ergebnis. Die reale Verteilung des stochastischen Signals wird zum ersten Mal von den beiden Kennlinien eingeschlossen und repräsentiert somit das anfänglich gesuchte Ergebnis<sup>123</sup>.

<sup>123</sup> vgl. Kapitel 7.3.1 Das Testsignal



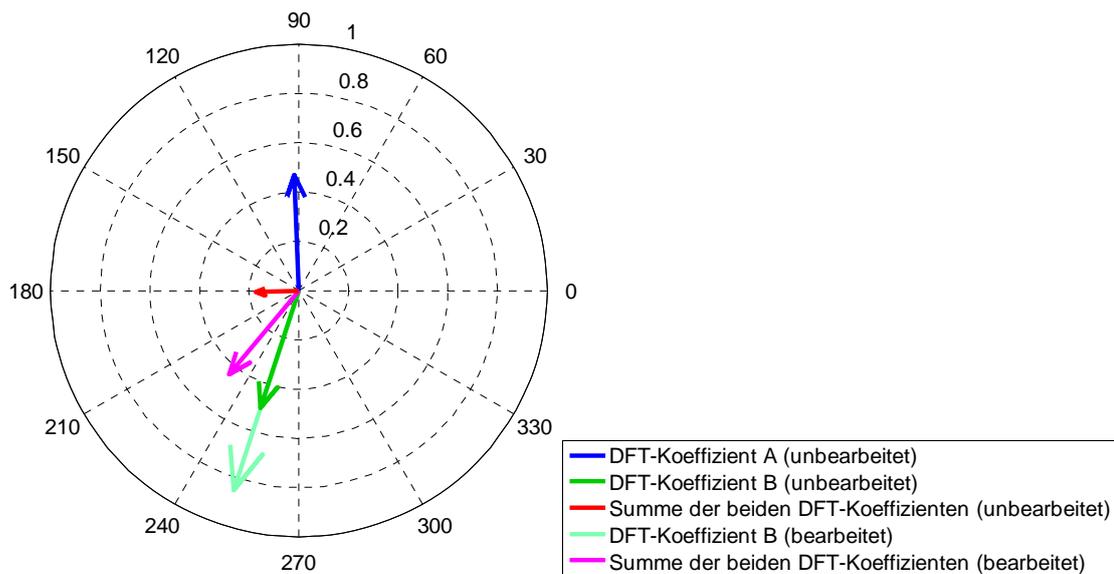
**Abbildung 7.3-11** Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix. Die Bearbeitung erfolgt mit einer Fallunterscheidung bei  $IST/SOLL = 1$ . Die Bearbeitung in Fall 1 ( $IST < SOLL$ ) erfolgt mit dem Korrekturfaktor  $x[p]$ , wobei  $p[k] = Re(A[k])Re(B[k]) + Im(A[k])Im(B[k])$ . Der DFT-Koeffizient mit der größeren Amplitude wird bearbeitet.

Der subjektive Eindruck hingegen ist ernüchternd. Zwar sind die Artefakte weniger stark wahrzunehmen als in Keils Version, im Vergleich zur Variante „IRT\_5i“ ist hingegen eine Verschlechterung auszumachen.

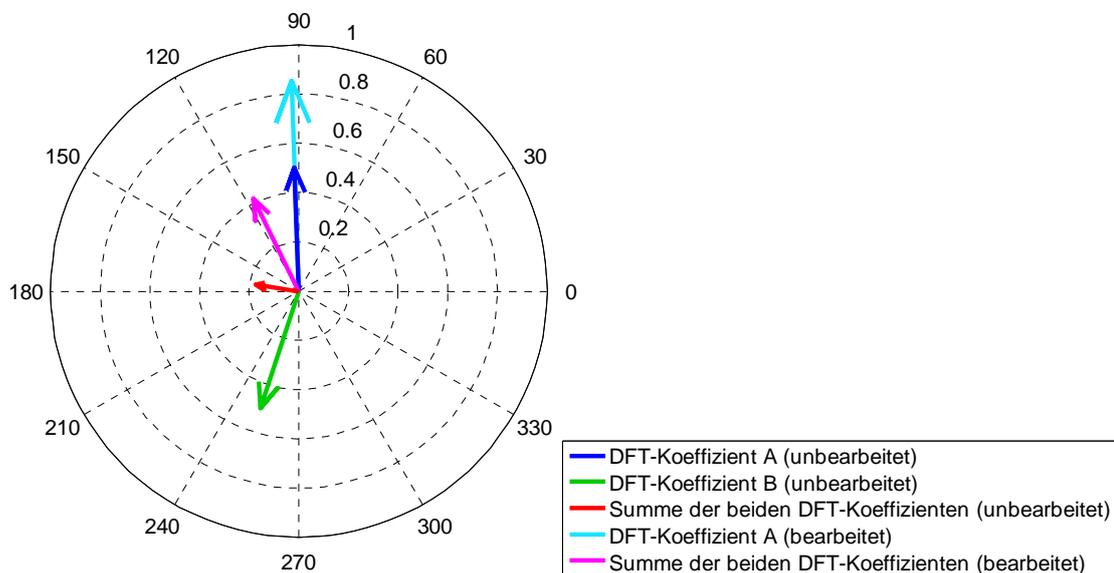
Grund für die Artefakte sind Phasensprünge, welche durch die wechselnde Bearbeitung des ersten und zweiten DFT-Koeffizienten herbeigeführt werden.

Es sei angenommen, dass ein zu summierendes DFT-Koeffizienten-Paar einen sehr geringen Größenunterschied bei seinen Amplituden aufweist. Die Wahl, welcher Koeffizient bearbeitet wird, würde damit denkbar knapp ausfallen. Hinsichtlich der Amplitude der bearbeiteten Summe ist diese Wahl unwesentlich. Jedoch unterscheidet sich die Phasenlage, je nachdem welcher Koeffizient bearbeitet wird, erheblich. Verändern sich die Amplituden der beiden zu summierenden Koeffizienten von Fenster zu Fenster derart, dass einmal der erste und im darauf folgenden Fenster der zweite Koeffizient bearbeitet wird, entstehen Phasensprünge, die wahrnehmbare Knackgeräusche auslösen können.

Abbildung 7.3-12 und Abbildung 7.3-13 stellen zwei nahezu identische Ausgangssituationen dar. Lediglich die Amplitude von  $B[r]$  ist in der ersten Abbildung minimal größer als die von  $A[r]$ . In der zweiten Abbildung verhält sich das Größenverhältnis andersherum. In Bezug auf die Phase der bearbeiteten Summe zeigen die beiden Abbildungen zwei völlig unterschiedliche Ergebnisse.



**Abbildung 7.3-12** Darstellung der an einem Additionsvorgang des IRT-Downmixes beteiligten DFT-Koeffizienten in der komplexen Ebene. Der Algorithmus bearbeitet den DFT-Koeffizienten mit der größeren Amplitude. Der Phasenwinkel wird radial, die Amplitude zentrifugal dargestellt.  $A[r] = 0,48 \cdot e^{1,59 \cdot i}$  und  $B[r] = 0,51 \cdot e^{4,39 \cdot i}$ .



**Abbildung 7.3-13** Darstellung der an einem Additionsvorgang des IRT-Downmixes beteiligten DFT-Koeffizienten in der komplexen Ebene. Der Algorithmus bearbeitet den DFT-Koeffizienten mit der größeren Amplitude. Der Phasenwinkel wird radial, die Amplitude zentrifugal dargestellt.  $A[r] = 0,51 \cdot e^{1,59 \cdot i}$  und  $B[r] = 0,48 \cdot e^{4,39 \cdot i}$ .

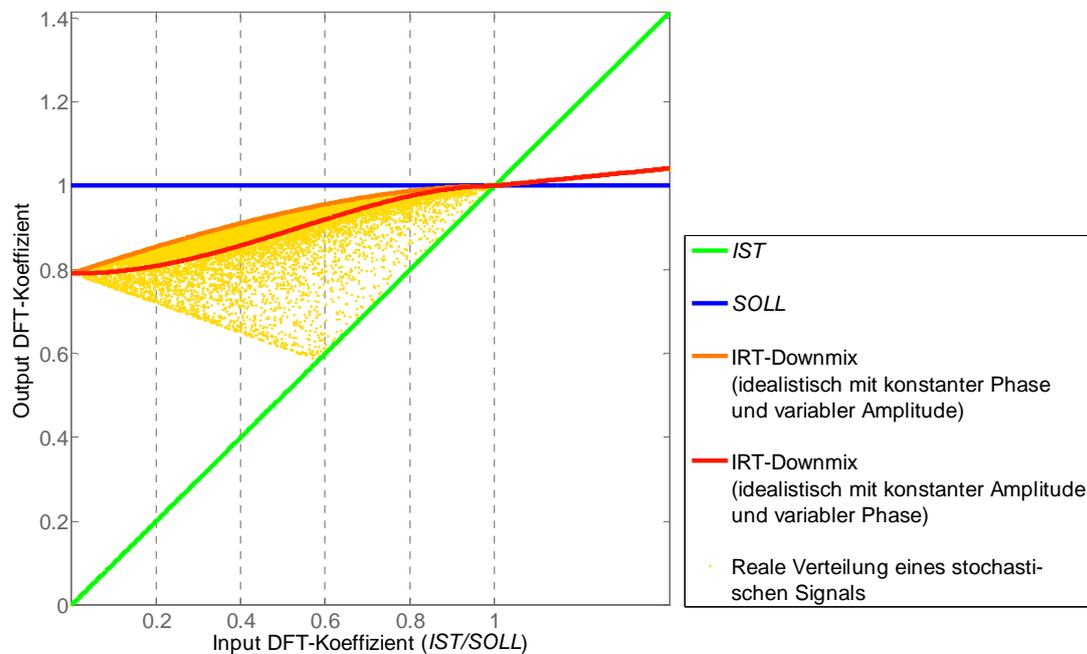
Die Summe zweier sich vollständig auslöschender Signale besitzt auch keine Information über ihren Phasenwinkel. Durch die dynamische Auswahl eines der beiden zu summierenden DFT-Koeffizienten für die Bearbeitung ist der resultierende Phasenwinkel uneindeutig.

### 7.3.5 Lösungsansätze

Die bisherigen Untersuchungen haben gezeigt, dass neben einer unstetigen Kennlinie vor allem ein uneindeutiger Phasenwinkel die Artefakte auslösen kann. Gute Ergebnisse lieferten bisher die Lösungsansätze, mit denen für jeden DFT-Koeffizient der Phasenwinkel eindeutig bestimmt werden kann, wie zum Beispiel der Algorithmus „IRT\_5i“. Dieser stellt daher die Basis für die weitere Optimierung dar.

#### Lösung der Dämpfungsproblematik

Die im Folgenden beschriebene, neue Variante sieht eine pragmatische Lösung des Dämpfungsproblems vor. Alle DFT-Koeffizienten, die nach der Bearbeitung im Fall 1 ( $0 < IST/SOLL < 1$ ) gegenüber dem *IST*-Wert gedämpft sind – also alle Punkte der realen Verteilung, die unter der grünen *IST*-Linie liegen – werden durch den zugehörigen *IST*-Wert ausgetauscht. Die Kennlinie der *IST*-Werte schließt damit die reale Verteilung nach unten hin ab und verhindert eine Verschlechterung in Folge einer Dämpfung (Abbildung 7.3-14).



**Abbildung 7.3-14** Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix „5ib“. Die Bearbeitung erfolgt mit einer Fallunterscheidung. Die Korrektur in Fall 1 ( $IST < SOLL$ ) erfolgt mit dem Korrekturfaktor  $x[p]$  aus (7.12), wobei  $w = 0,25$ . Für DFT-Koeffizienten, die mit der Bearbeitung eine Dämpfung erfahren, wird der zugehörige *IST*-Wert verwendet.

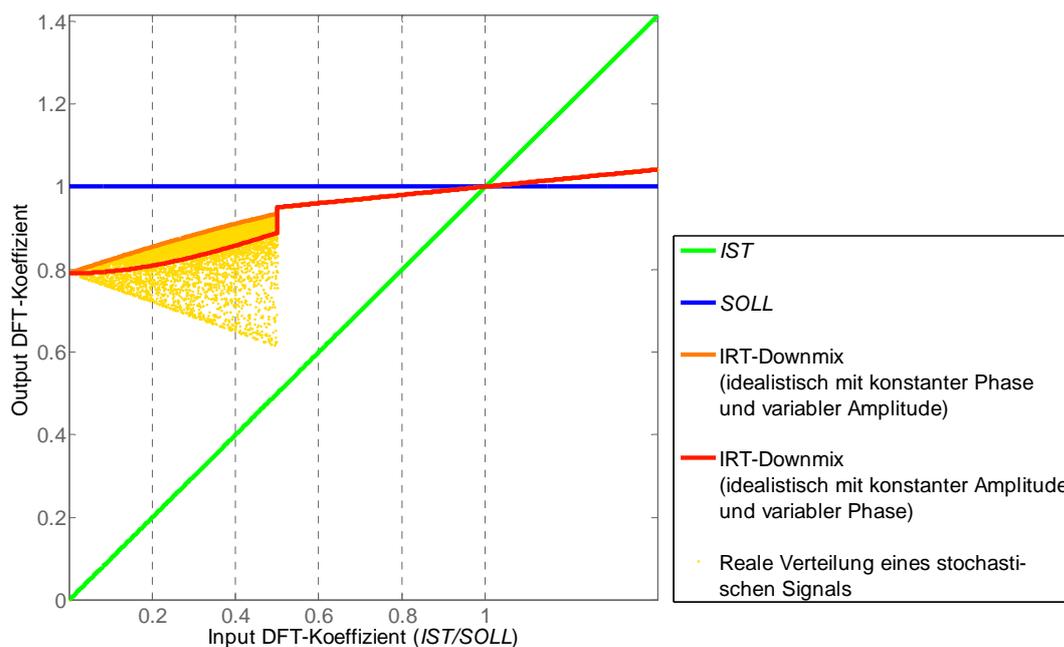
Abbildung 7.3-14 zeigt den erhofften Erfolg bei den Werten des stochastischen Signals. Die subjektive Analyse ist ebenfalls sehr vielversprechend und offenbart eine noch ausgewogenere Klangfarbe.

Damit ist, mit praktisch nicht wahrnehmbaren Störungen und einer ausgeglichenen Klangfarbe, eine schöne, erste Lösung gefunden. Diese Variante des IRT-Downmix erhält die Bezeichnung „IRT\_5ib“.

### Verschiebung der Position der Fallunterscheidung

Betrachtet man den kompletten Additionsprozess mit seinen mittlerweile nur noch zwei Fällen und einer Fallunterscheidung bei  $IST/SOLL = 1$ , so fällt auf, dass der problematische Bereich der des ersten Falls ist. Die zu summierenden DFT-Koeffizienten sind nur schwer und mit subjektiv wahrnehmbaren Einbußen derart zu bearbeiten, dass sich ein einheitliches Verstärkungsverhältnis, also eine für alle Werte gültige Kennlinie, ergibt.

Im Gegensatz dazu hat die Bearbeitung der Summe in Fall 2 keinerlei negative Folgen. Die Phasenlage ist eindeutig und die Verstärkung proportional zum Korrekturfaktor. Daher ist es durchaus erwägenswert, den Bereich für Fall 2 zu vergrößern und so die problematische Bearbeitung in Fall 1 einzuschränken. Letztlich sind nur die zu summierenden DFT-Koeffizienten-Paare auf eine Bearbeitung vor der Addition, wie sie in Fall 1 angewandt wird, angewiesen, deren Summe vollständig oder nahezu ausgelöscht wird. Der Bereich zwischen  $IST/SOLL = 0,5$  und  $IST/SOLL = 1$  kann somit problemlos von Fall 2 bearbeitet werden. Der Korrekturfaktor würde maximal die Wertigkeit 2 erreichen, um eine gegenüber dem  $SOLL$ -Wert um 50% abgesenkte Summe zu korrigieren.



**Abbildung 7.3-15** Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix. Die Bearbeitung erfolgt mit einer Fallunterscheidung bei  $IST/SOLL = 0,5$ . Die Bearbeitung in Fall 1 erfolgt mit dem Korrekturfaktor  $x[p]$ , wobei  $p[k] = Re(A[k])Re(B[k]) + Im(A[k])Im(B[k])$ .

Zwei Probleme wirft dieser Ansatz auf: Der Übergang von Fall 1 zu Fall 2 verläuft nicht immer stetig. Außerdem können Phasensprünge bei diesem Übergang nicht ausgeschlossen werden.

### Korrektur in zwei Durchläufen

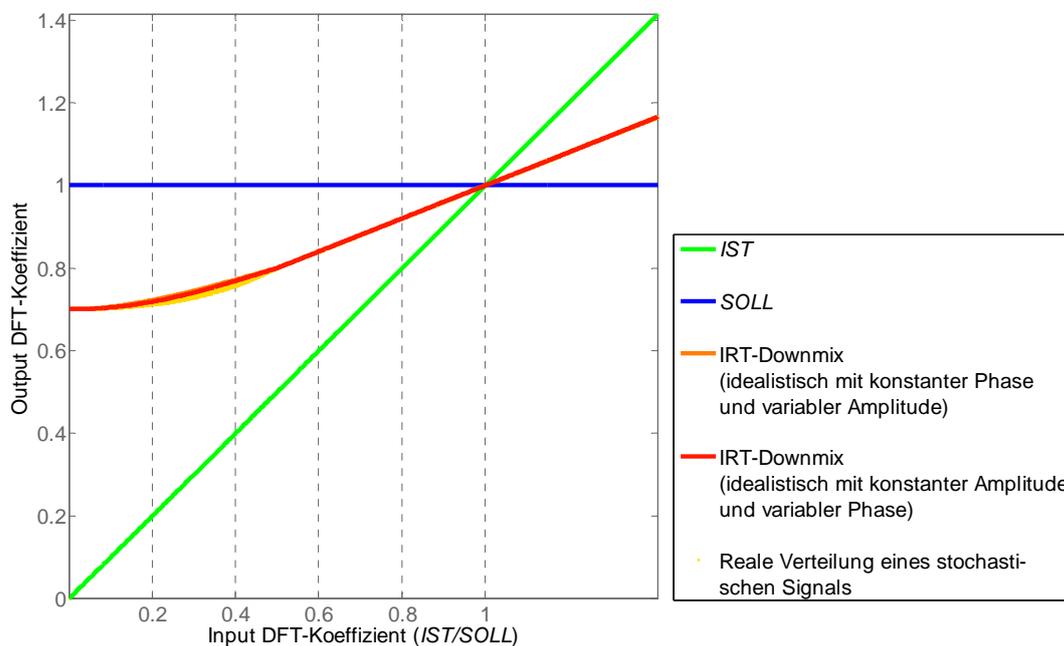
Wie bereits im vorangegangenen Kapitel angedeutet, ist die Bearbeitung eines der zu summierenden DFT-Koeffizienten im Fall 1 nur deshalb nötig, weil bei einer starken Absenkung oder einer

Auslöschung ein bis zu unendlich großer Korrekturfaktor nötig wäre, um die kleine oder überhaupt nicht vorhandene Amplitude der Summe zu kompensieren.<sup>124</sup>

Um in diesem Bereich ebenfalls in den Genuss der, gegenüber der Vorgehensweise in Fall 1 zu bevorzugenden Bearbeitung der Summen zu kommen, wird der Bearbeitungsprozess in zwei Durchläufe unterteilt. Zuerst wird durch die Bearbeitung des ersten zu summierenden DFT-Koeffizienten, die Summe auf ein gewisses Grundniveau angehoben. Eine einheitliche Kennlinie ist für diesen Durchlauf nicht notwendig. Lediglich die Anhebung auf etwa 50% des *SOLL*-Wertes soll eine weitere Bearbeitung der resultierenden Summe ermöglichen.

In einem zweiten Durchlauf werden die DFT-Koeffizienten der so entstandenen Summe auf ein einheitliches Verstärkungsverhältnis korrigiert, so dass eine für alle Werte gültige Kennlinie entsteht. Der Phasenwinkel wird wie in Fall 2 vom *IST*-Zustand übernommen.

Die Vorteile der beiden Vorgehensweisen in den Fällen 1 und 2 werden somit geschickt vereint. Die Bearbeitung des ersten zu summierenden DFT-Koeffizienten verhindert eine Auslöschung oder zu starke Absenkung. Durch die danach folgende Bearbeitung der Summe wird die Divergenz der Amplituden ausgeglichen. Der Übergang von Fall 1 zu Fall 2 ist jetzt direkt beeinflussbar und kann flexibel angepasst werden, um einen stetigen Verlauf der Kennlinie zu erzielen. Auch die Phase springt an dieser Stelle nicht mehr, da in beiden Fällen der Winkel des *IST*-Zustands herangezogen wird.



**Abbildung 7.3-16** Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix „5j“. Die Bearbeitung erfolgt mit einer Fallunterscheidung bei  $IST/SOLL = 0,5$ , und im 1. Fall in zwei Durchläufen. Die Bearbeitung in Fall 2 erfolgt bereits anhand der Optimierung in Kapitel 8.

<sup>124</sup> vgl. Kapitel 6.1.1 Kammfilterkompensation

Abbildung 7.3-16 zeigt das Ergebnis mit den beschriebenen Veränderungen. Der subjektive Hör-eindruck bestätigt weitgehend die theoretischen Überlegungen. Die klanglichen Eigenschaften sind vielversprechend und können als vergleichbar mit der Lösung „IRT\_5ib“ eingestuft werden. Bei dem Downmix eines künstlich generierten Signals, das einen starken Kammfilter erzeugt<sup>125</sup> offenbart diese Version jedoch einen unangenehmen, neuen Artefakt. Dieser an- und abschwel-lende Ton im oberen Frequenzbereich ist mit einem „Ringing“ zu vergleichen.

Die Variante erhält den Namen „IRT\_5j“.

---

<sup>125</sup> vgl. Kapitel 9.1.2 Hörbeispiel C: Kammfilter

## 8 Analyse und Optimierung des IRT-Downmixes hinsichtlich der Pegeländerung und Phantomschallquellenverschiebung

Nachdem der Algorithmus auf die Kompensation von Kammfiltern und Differenzpegeländerungen hin optimiert wurde, soll abschließend noch die beim Downmix entstehende Pegeländerung und Abbildungsverschiebung von Phantomschallquellen untersucht und verbessert werden.

### 8.1 Analyse

Für die objektive Analyse werden mehrkanalige Testsignale mit Phantomschallquellen im Frontbereich zwischen L und C künstlich generierten. Diese werden mit den verschiedenen Downmixverfahren herunter gemischt und die Ergebnisse hinsichtlich der im Stereo-Signal entsprechend entstandenen Phantomschallquellen ausgewertet.

Bei der Erzeugung des Testsignals ist es wichtig, dass alle Phantomschallquellen das gleiche Energieniveau beim Hörer erreichen. Die resultierenden Pegel bei der Wiedergabe des Mehrkanaltons  $P_{5,0}$  und bei der Wiedergabe des Downmixes  $P_{2,0}$  können mit (4.8) ermittelt werden:

$$P_{5,0} = 10 \cdot \log_{10} \left( 10^{\frac{P_{5,0L}}{10}} + 10^{\frac{P_{5,0C}}{10}} \right), \quad (8.1)$$

$$P_{2,0} = 10 \cdot \log_{10} \left( 10^{\frac{P_{2,0L}}{10}} + 10^{\frac{P_{2,0R}}{10}} \right). \quad (8.2)$$

Der Auslenkungswinkel  $\alpha_{5,0}$  einer Phantomschallquelle zwischen L und C in der Mehrkanalversion bzw.  $\alpha_{2,0}$  in der Downmixversion kann mit Hilfe der Abbildung 4.5-1 wie folgt definiert werden:<sup>126</sup>

$$\alpha_{5,0} = 15^\circ + \Delta P_{5,0} \cdot 1,1^\circ, \quad \text{wobei } \Delta P_{5,0} = P_{5,0L} - P_{5,0C}, \quad (8.3)$$

$$\alpha_{2,0} = \Delta P_{2,0} \cdot 2,2^\circ, \quad \text{wobei } \Delta P_{2,0} = P_{2,0L} - P_{2,0R}. \quad (8.4)$$

Setzt man (8.3) in (8.1) ein und löst das Gleichungssystem nach  $P_{5,0C}$  auf, so erhält man:<sup>127</sup>

$$P_{5,0C} = 10 \cdot \log_{10} \left( \frac{10^{\frac{P_{5,0}}{10}}}{10^{\frac{\alpha_{5,0} - 15^\circ}{1,1^\circ} \cdot \frac{P_{5,0L}}{10}} + 1} \right). \quad (8.5)$$

Mit (8.5) lässt sich für einen konstanten Gesamtpegel  $P_{5,0}$  in Abhängigkeit des Auslenkungswinkels  $\alpha_{5,0}$  der notwendige Pegel für das Centersignal  $P_{5,0C}$  bestimmen. Der Pegel für das linke Signal  $P_{5,0L}$  kann schließlich durch Umformung von (8.1) berechnet werden:

$$P_{5,0L} = 10 \cdot \log_{10} \left( 10^{\frac{P_{5,0}}{10}} - 10^{\frac{P_{5,0C}}{10}} \right) \quad (8.6)$$

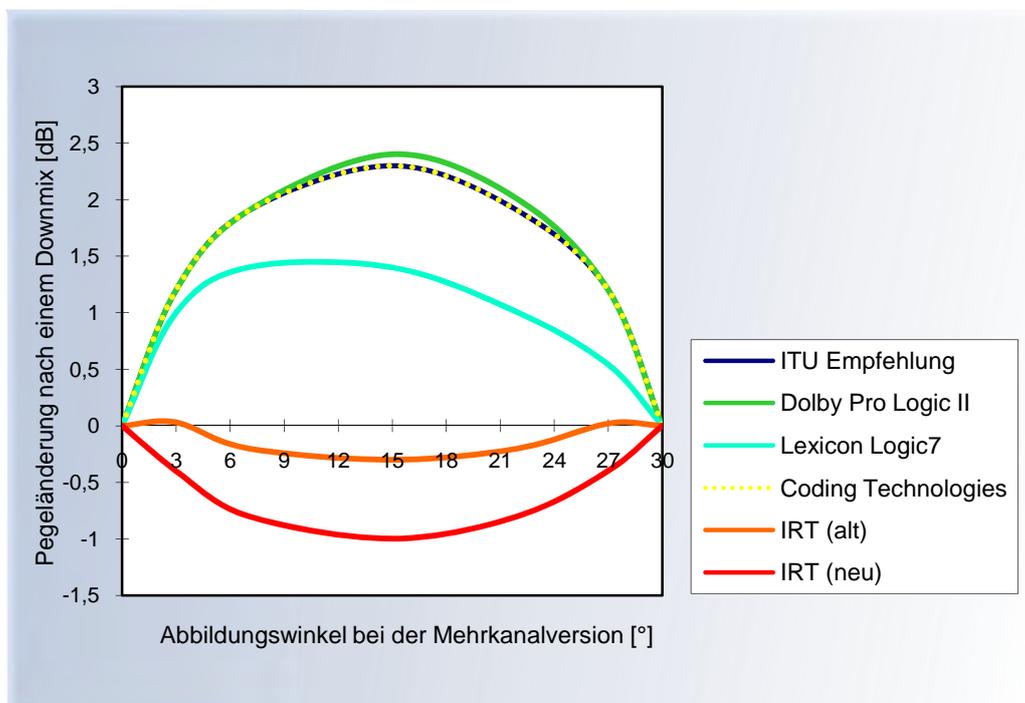
Auf diese Weise lassen sich gleichwertige Phantomschallquellen zwischen L und C generieren, und nach dem Downmix hinsichtlich ihres Pegels und des Abbildungswinkels mit der Mehrkanalversion vergleichen.

<sup>126</sup> vgl. Kapitel 4.5.1 Pegeldifferenzstereofonie

<sup>127</sup> Der vollständige Lösungsweg kann im Anhang auf Seite 125 nachvollzogen werden.

### 8.1.1 Pegeländerung

Abbildung 8.1-1 zeigt die Pegeländerung der einzelnen Verfahren nach dem Downmix. Aus dem orangefarbenen Graphen des Downmixverfahren „IRT (alt)“ wird ersichtlich, dass die Kammfilterkompensation wie vermutet<sup>128</sup> auch die vom Abbildungswinkel abhängigen Pegeländerungen nachhaltig korrigiert.



**Abbildung 8.1-1** Pegelzunahme nach einem Downmix in Abhängigkeit vom Abbildungswinkel einer Phantomschallquelle in der Mehrkanalversion.

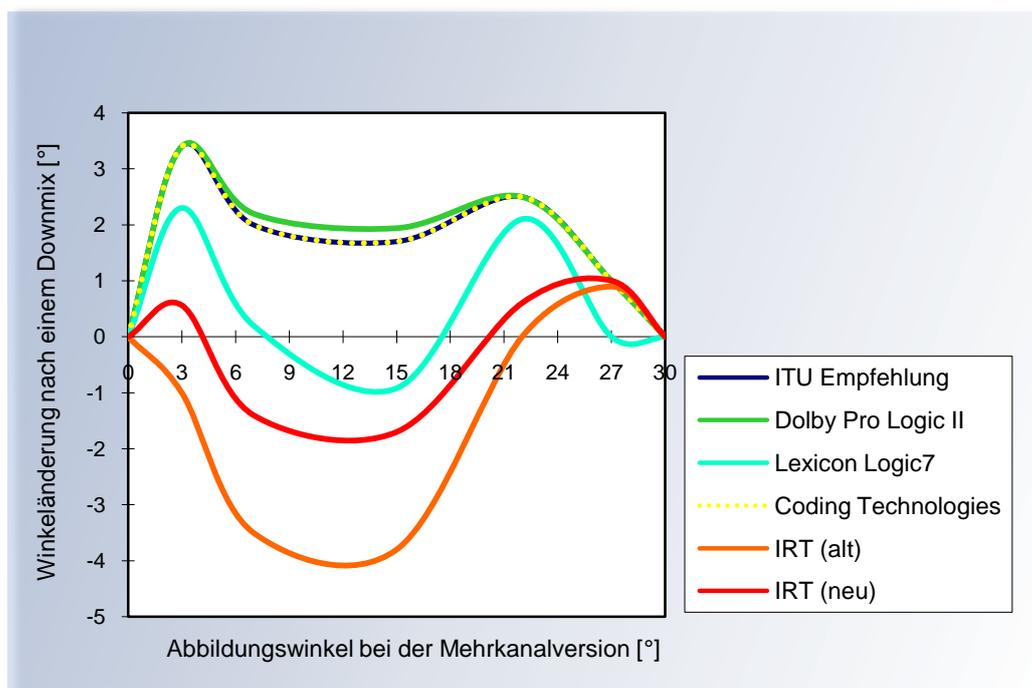
Eine maximale Änderung von 0,3 dB ist im Vergleich zu den anderen Verfahren ein sehr gutes Ergebnis, das keiner weiteren Optimierung bedarf. Auch subjektiv kann dieses Resultat bestätigt werden.

### 8.1.2 Phantomschallquellenverschiebung

Aus Abbildung 8.1-2 kann die Winkeländerung einer Phantomschallquelle in Abhängigkeit des Abbildungswinkels bei der Mehrkanalwiedergabe entnommen werden. Das erhoffte Ergebnis, dass auch die Verschiebung durch die Kammfilterkompensation korrigiert wird, bleibt aus. Vergleicht man den orangefarbenen Graphen des IRT-Downmixes (alt) mit den anderen, insbesondere mit dem von Logic7, dessen Algorithmus ebenfalls eine Kompensation der Phantomschallquellenverschiebung vornimmt<sup>129</sup>, so ist eine ähnlich Form auszumachen, die jedoch in den negativen Bereich verschoben ist. Dies deutet - im Gegensatz zur Pegeländerung - auf eine übermäßige Korrektur hin, welche, das Optimum überschreitend, bereits zum anderen Extrem führt.

<sup>128</sup> vgl. Kapitel 6.1.5 Eine Lösung für vier Probleme

<sup>129</sup> vgl. Kapitel 3.3.1 Lexicon Logic7



**Abbildung 8.1-2** Änderung des Abbildungswinkels einer Phantomschallquelle nach einem Downmix in Bezug auf die Mehrkanalversion. Ein positives Vorzeichen entspricht einer Verschiebung nach außen, ein negatives Vorzeichen einer Verschiebung nach innen zur Mitte hin.

Träfe diese Vermutung zu, würde sie im Umkehrschluss bedeuten, dass sich eine vollständige Korrektur der Änderung des Abbildungswinkels und des Pegels einer Phantomschallquelle nach dem Downmix ausschließen. In einem solchen Fall kann die Lösung nur auf einem Kompromiss beruhen, der nach subjektiven Kriterien ermittelt wird.

## 8.2 Korrektur

Die Veränderungen einer Phantomschallquelle nach dem Downmix in Abbildung und Pegel sind auf die unterschiedlichen Additionseigenschaften von elektrischen Signalen im Vergleich zu akustischen Signalen zurückzuführen.<sup>130</sup> Die um bis zu 3 dB größere elektrische Summe zweier kohärenter Signale gegenüber der akustischen Addition führt bei einem Downmix zu einer Anhebung im Vergleich zum *SOLL*-Wert. Der DFT-Koeffizient einer solchen Summe liegt in dem Kennlinien-Diagramm im Bereich:  $1 < IST/SOLL < 1,4142$  (Abbildung 7.3-14). Auf ihn trifft also Fall 2 zu (bzw. in Keils Version Fall 4), bei dem die Summe bearbeitet wird. Die Kennlinie (Abbildung 7.3-14) weist in diesem Bereich tatsächlich einen sehr großen Kompensationsgrad auf. Die Abweichung vom *SOLL*-Wert beträgt nur 10% der Differenz von *IST*- und *SOLL*-Wert.<sup>131</sup>

Versuche mit einem geringeren Kompensationsgrad in Fall 2 bestätigten die Vermutung aus dem vorangegangenen Kapitel, dass sich die Korrektur von Pegel- und Verschiebungsfehler nicht vollständig vereinen lassen. Nach weiteren Testreihen und unter Berücksichtigung der Ergebnisse von

<sup>130</sup> vgl. 5.2 Veränderung der effektiven Energie sowie 5.4 Phantomschallquellenverschiebung

<sup>131</sup>  $|S_{bearbeitet}| = SOLL + (IST - SOLL) \cdot 0,1$ . Vgl. Kapitel 6.2.2 Der Algorithmus

Deigmöllers Untersuchungen<sup>132</sup> fiel die Entscheidung auf eine Kompromisslösung, die eine Abweichung der bearbeiteten Summe gegenüber dem *SOLL*-Wert von 40% der Anhebung vorsieht.

$$|S_{bearbeitet}| = SOLL + (IST - SOLL) \cdot 0,4. \quad (8.7)$$

Die dadurch resultierenden Einbußen bei der Pegeländerung im Bereich von Anhebungen liegen bei maximal 1 dB. Hält man sich vor Augen, dass Anhebungen bei einem Kammfilter mit einer Pegeländerung von maximal 3 dB weniger Einfluss auf die Veränderung der Klangfarbe nehmen als Absenkungen, ist dieser Wert durchaus zu vertreten. Auf die Kompensation der Absenkungen oder gar Auslöschungen hat diese Veränderung keinen Einfluss, da diese Fall 1 zugewiesen werden.

Die roten Graphen zeigen in Abbildung 8.1-1 und Abbildung 8.1-2 den IRT-Downmix (neu) mit den beschriebenen Veränderungen.

---

<sup>132</sup> vgl. Kapitel 3.4.1 Ergebnis des Hörversuchs sowie 6.1 Wesentliche Ansatzpunkte

## 9 Hörversuch

### 9.1 Durchführung

Zur Evaluation des im Rahmen dieser Arbeit optimierten IRT-Downmixes wird dieser in einem Hörversuch mit anderen Verfahren und einer von Hand erstellten Mischung verglichen. Dabei sollte die im Vorfeld des Versuchs favorisierte Variante des IRT-Downmixes „5ib“ sowohl der Variante „5keil“ von Martin Keil, der Version „5j“, einem Downmix nach der ITU-Empfehlung als auch dem zugehörigen Handmix gegenüber gestellt werden. Außerdem wurde der Vergleich der Variante „5j“ mit der ITU-Empfehlung in den Hörversuch aufgenommen.

An der Evaluation nahmen 12 Mitarbeiter des Instituts für Rundfunktechnik und des Bayerischen Rundfunks teil. Alle Teilnehmer sind als Experten anzusehen.

Als Versuchsmethode wurde ein Paarvergleich (A-B-Vergleich) gewählt, da weder der Mehrkanalton noch eine der Stereoverversionen als Referenz angesehen werden kann. Dem Versuchsteilnehmer standen damit zwei Stereoverversionen A und B gegenüber, die er hinsichtlich ihrer subjektiv wahrgenommenen Qualität vergleichen sollte. Subjektive Kriterien können beispielsweise die Klangfarbe, Lautstärkebalance, Räumlichkeit, Transparenz, Abbildung oder die Artefaktbildung sein.

Die Durchführung des Hörversuchs wurde mit dem Programm „STEP“ realisiert. Über einen kleinen Monitor konnten die Versuchsteilnehmer das Programm mit Hilfe einer Maus oder Tastatur bedienen. Mit den Tasten „A“ und „B“ konnte die wiederzugebende Stereoverversion ausgewählt werden. Die Taste „REF“ aktivierte den ursprünglichen Mehrkanalton, der jedoch ausschließlich zur Orientierung verwendet werden sollte. Auch ein Wechsel der Version während der Wiedergabe war möglich. Mit den Schiebereglern „Start“ und „Stop“ konnte der wiederzugebende Bereich eingeschränkt und mit dem Schieberegler „Position“ die Abspielposition bestimmt werden. Über die grafisch zu interpretierenden Tasten für „Play“ und „Pause“ konnte die Wiedergabe gestartet und angehalten werden. Die Taste „Loop“ aktivierte die Wiederholungsfunktion bei der Wiedergabe. Im oberen Teil des Programms war es dem Versuchsteilnehmer schließlich möglich, seine Bewertung abzugeben. In einer siebenstufigen Skala konnten die Wertungen von „A is much better than B“ bis „B is much better than A“ ausgewählt werden (Abbildung 9.1-1). Mit Hilfe der Taste „Next“ konnte der Teilnehmer zum nächsten Vergleich gelangen.

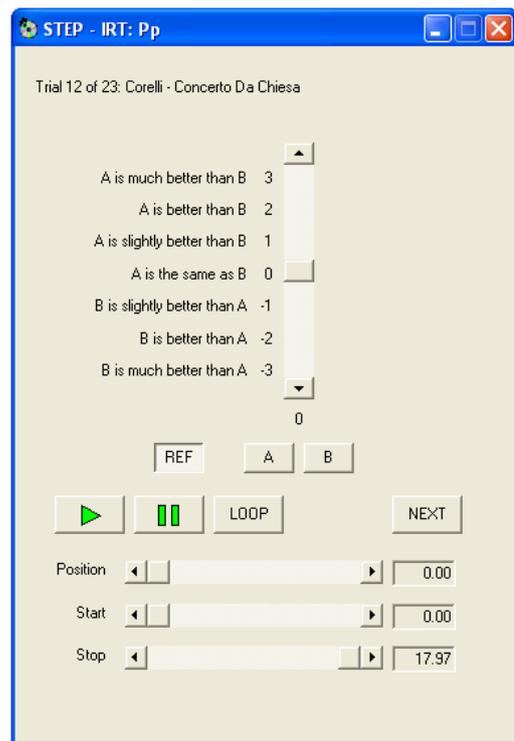


Abbildung 9.1-1 Oberfläche des Programms „STEP“

Jeder Versuchsteilnehmer wurde nach einer schriftlichen Einweisung<sup>133</sup> mit einer Trainingssession in den Ablauf des Versuchs sowie die Bedienung des Programms eingeführt.

Während des Hörversuchs saßen die Teilnehmer im Sweet-Spot der Lautsprecheraufstellung.

### 9.1.1 Versuchsaufbau

Die Durchführung des Hörversuchs fand im Tonstudio des Instituts für Rundfunktechnik statt. Die digitalen Ausgangssignale des Programms „STEP“ wurden von der RME-Soundkarte HDSP AES-32 ausgegeben und über eine D.A.I.S.-Digitalmatrix und das Digital Mischpult DM2000 der Firma Yamaha an einen D/A-Wandler (Modell 205) der Firma Jünger weitergeleitet. Die Wiedergabe erfolgte über fünf Studiolautsprecher vom Typ „RL922K“ der Firma Geithain. Jeder dieser Studiolautsprecher wurde anhand der Empfehlung „Hörbedingungen und Wiedergabeanordnungen für Mehrkanalstereofonie“ des Surround Sound Forums eingemessen.<sup>134</sup>

### 9.1.2 Auswahl des Testmaterials

Die Auswahl des Testmaterials sollte neben einem breit gefächerten Spektrum unterschiedlicher, in der Praxis auftretender Genres auch solches Material berücksichtigen, das verschiedene, für einen Downmix kritische Situationen hervorruft. Daher wurden neben neun praxisnahen Hörbeispielen auch drei synthetisch erzeugte Signale ausgewählt. Die insgesamt zwölf Hörbeispiele wurden in drei Sessions à 20 Minuten aufgeteilt.

#### Hörbeispiel 1: Chor

Der verwendete Ausschnitt stammt aus der „Missa Papae Marcelli“ von Giovanni Pierluigi da Palestrina. Die Aufnahme ist eine Liveproduktion des Bayrischen Rundfunks und entstand bei einem Konzert im Vatikan.

Grund für die Aufnahme dieser Sequenz in den Hörversuch ist die starke Artefaktbildung beim Downmix mit dem Verfahren von Keil. Außerdem bietet die Anordnung der Chorsänger die Möglichkeit zur Beurteilung der Abbildungsstabilität der Downmixverfahren.

#### Hörbeispiel 2: Klassik1

Das zweite Hörbeispiel ist eine Sequenz aus dem Allegro des „Concerto da chiesa“ von Arcangelo Corelli. Die Aufnahme entstand mit der New Dutch Academy.

Diese aufwendige Klassikproduktion bietet die Möglichkeit zur Beurteilung der Klangfarbe, Abbildungsstabilität und Räumlichkeit nach einem Downmix.

#### Hörbeispiel 3: TV-Show

Ein Livemitschnitt der Fernsehshow „Stars in der Manege“ aus dem Jahr 2006 bringt neben dem kritischen Applausignal typische Showelemente in den Hörversuch mit ein. Interessant sind die Off-Air-Moderationsstimme, die in der Mehrkanalversion über alle Lautsprecherkanäle wiedergegeben wird, und die mehrkanalig abgemischte Showband.

---

<sup>133</sup> vgl. hierzu Anhang S. 130

<sup>134</sup> vgl. Theile & Krämer 2002

### Hörbeispiel 4: Hörspiel

Der Hörspielausschnitt stammt aus der WDR-Produktion „Das Herz von Jane Mansfield“ von Armin Frost. Neben der im Frontbereich rechts und links angeordneten männlichen und weiblichen Stimme bieten Geräusche in der kompletten Horizontalebene und ein Sprecher Möglichkeiten zur Beurteilung der Abbildungsstabilität und vor allem der Lautstärkebalance nach dem Downmix.

### Hörbeispiel 5: Klassik2

Diese elfsekündige Sequenz ist dem Violinkonzert in d-moll Op. 47 von Jean Sibelius entnommen. Die Aufnahme wurde vom Bayerischen Rundfunk unter Live-Bedingungen im Herkulesaal der Residenz in München erstellt.

Auffällig beim Downmix ist die starke Bildung von Kammfiltereffekten, die durch eine ungünstige Mikrofonierung entstehen. Daher eignet sich dieses Hörbeispiel sehr gut zur Beurteilung einer Änderung der Klangfarbe.

### Hörbeispiel 6: Jazz1

Diese Aufnahme des Bayerischen Rundfunks ist im Rahmen der Veranstaltung „Jazz Burghausen“ mit dem Lynne Arriale Trio im Jahr 2005 entstanden. Die Mikrofonierung erfolgte mit dem Ziel, Phasenprobleme zwischen den fünf Kanälen zu beseitigen. Ein Downmix nach der ITU-Empfehlung sollte somit bereits zu guten Ergebnissen führen.

### Hörbeispiel 7: Jazz2

Das zweite Jazz-Beispiel entstammt im Gegensatz zu Hörbeispiel 6 einer Studioaufnahme. Der Titel „Lament for Linus“ wurde mit der Band „ACF“ (Advanced Combo Funk) an der Hochschule der Medien in Stuttgart unter der Betreuung von Prof. Oliver Curdt produziert. Die Instrumente wurden einzeln mikrofoniert und teilweise im Overdub-Verfahren<sup>135</sup> aufgenommen. Die Mehrkanalmischung erstellte Dennis Kopac.

Die Anordnung des Fagotts, der Trompete und der Talking-Drum bietet Möglichkeiten zur Beurteilung der Abbildungsstabilität und Räumlichkeit. Interessant sind auch das Fender-Rhodes, das in der Mehrkanalversion größtenteils über die Surroundkanäle wiedergegeben wird, und das allmählich einsetzende Schlagzeug, das ausschließlich auf dem Center-Kanal liegt.

### Hörbeispiel 8: Orgel

Diese Sequenz ist von der SACD „Swedish Folk Tunes“ des schwedischen Duos Latar. Die Aufnahme einer Orgel mit einer Nyckelharpa in einem komplexen Raum, wie einer Kirche, bietet die Möglichkeit zur Beurteilung der Abbildung, Räumlichkeit und Klangfarbe. Der flächige Klang der Orgel stellt hinzukommend eine Herausforderung für Downmixverfahren dar.

### Hörbeispiel 9: Pop

Für die Beurteilung des Downmixes von Pop-Musik wurde der Titel „My Love Don't Cost A Thing“ von Jennifer Lopez verwendet. Interessant sind hier die zusätzlichen Sound-Elemente der Mehr-

---

<sup>135</sup> Overdub ist eine aus dem englischen Sprachraum übernommene Bezeichnung für die Tonaufnahme über bereits vorhandene Aufnahmen im Tonstudio. Vgl. Overdub 2008.

kanalversion im Vergleich zur Stereoverversion. Dadurch resultierende Verdeckungseffekte oder ein überladenes Klangbild können sich bei einem automatischen Downmix nachteilig auswirken.

### **Hörbeispiel A: Intensitätsstereofonie**

Um ausschließlich die Abbildungsstabilität einer durch Pegeldifferenzen erzeugten Phantom-schallquelle bewerten zu können, wird in diesem Hörbeispiel ein Monosignal einer Kastagnette mit Hilfe der Intensitätsstereofonie (Pegeldifferenzstereofonie) vom linken Lautsprecher über den Centerkanal bis zum rechten Lautsprecher gepannt.

### **Hörbeispiel B: Laufzeitstereofonie**

Zur ausschließlichen Bewertungen der Laufzeitstereofonie wird das Signal der Kastagnette in dieser Sequenz mittel Laufzeiten vom linken über den Center- bis zum rechten Kanal gepannt.

### **Hörbeispiel C: Kammfilter**

Die Bewertung der Kammfilterkompensation wird durch einen künstlich provozierten Kammfiltereffekt möglich. Ein einkanaliger Signalausschnitt des Titels „Tom’s Diner“ von Suzanne Vega wird zeitversetzt auf die Kanäle L, C und R gelegt, so dass beim Downmix dieser klangverfälschende Filtereffekt auftritt.

## **9.1.3 Statistische Auswertung**

Die statistische Auswertung ist in erster Linie abhängig von der Art der Versuchsdurchführung.

Die verwendete Bewertungsskala ist in die Klasse der Ordinalskalen einzuordnen. Diese weisen keine linearen Abstände zwischen den Bewertungsmöglichkeiten auf. Zum Beispiel kann nicht ausgesagt werden, dass die Einschätzung „A is better than B“ der doppelt so guten Bewertung entspricht, wie „A is slightly better than B“. Die Größe der Bereiche zwischen den einzelnen Bewertungsmöglichkeiten kann nicht definiert werden.

Wenn eine Ordinalskala vorliegt, eignen sich für die grafische Darstellung in einem Diagramm der Medianwert und die Quartile.

Der Medianwert ist bei einer ungeraden Anzahl von Bewertungen genau der Wert, der bei einer nach der Größe geordneten Rangliste genau in der Mitte liegt. Liegt eine gerade Anzahl von Bewertungen vor, wie in dem vorliegenden Hörversuch, so wird der Mittelwert aus den beiden mittleren Bewertungen gebildet - in unserem Fall des sechsten und siebten Werts.

Quartile teilen diese, der Größe nach geordnete Rangliste in vier gleichgroße Gruppen ein. Das untere Quartil beschreibt den Punkt in der Rangliste, unter dem 25% aller Werte liegen. Das obere Quartil beschreibt analog dazu den Punkt, über dem 25% aller Werte liegen.

Die Auswertung mit Medianwert und Quartilen eignet sich besonders bei subjektiven Bewertungen. Ein sich von der Masse der Bewertungen stark unterscheidender Wert beeinflusst den Medianwert beispielsweise nicht so sehr wie das arithmetische Mittel.

## 9.2 Auswertung der Ergebnisse

Die folgenden Diagramme zeigen jeweils die Ergebnisse eines Vergleichspaares aller Hörbeispiele. Dabei wird der optimierte und favorisierte Algorithmus „IRT\_5ib“ im Vergleich zu den anderen Stereoersionen dargestellt.<sup>136</sup> Schneidet dieses neue Verfahren des IRT schlechter ab als sein Konkurrent, wird dies durch einen negativen Wert im Diagramm vermittelt. Schneidet die „5ib“-Version besser ab, liegt die Auswertung im positiven Wertebereich.

### 9.2.1 Vergleich IRT-Downmix „5ib“ mit der Downmixversion nach Keil

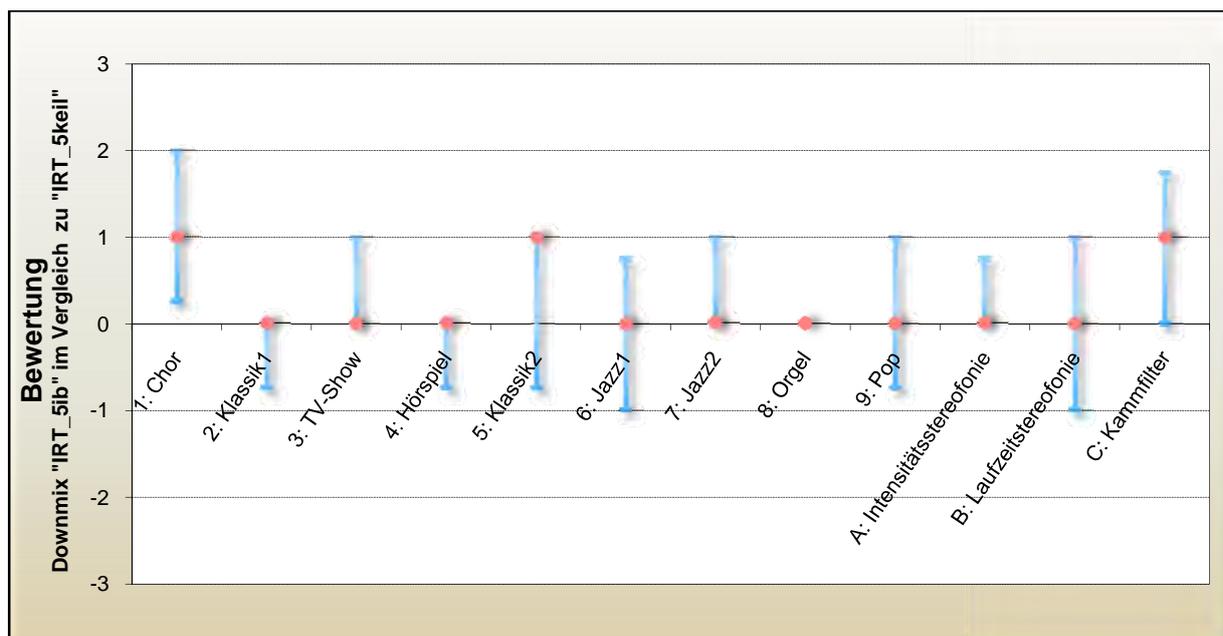


Abbildung 9.2-1 Downmix "IRT\_5ib" im Vergleich zu "IRT\_5keil"

Abbildung 9.2-1 zeigt die Erfolge der Optimierung, die im Rahmen dieser Arbeit entstanden sind. Deutlich zu erkennen ist die gute Bewertung beim ersten Hörbeispiel, bei welchem in Keils Version (IRT\_5keil) störende Artefakte entstehen. Die bessere Bewertung beim Hörbeispiel C und dem zweiten Klassikbeispiel, das beim Downmix ebenfalls besonders anfällig für Änderungen in der Klangfarbe ist, deutet auf eine gleichmäßigere Kammfilterkompensation und dadurch auf eine ausgewogenere Klangfarbe hin.

Die übrigen Hörbeispiele wurden als vergleichbar bewertet, so dass eine subjektiv wahrnehmbare Verschlechterung durch die Optimierung ausgeschlossen werden kann.

<sup>136</sup> Mit Ausnahme des Vergleichs der Variante „IRT\_5j“ mit der ITU-Empfehlung.

## 9.2.2 Vergleich IRT-Downmix „5ib“ mit dem Downmix nach der ITU-Empfehlung

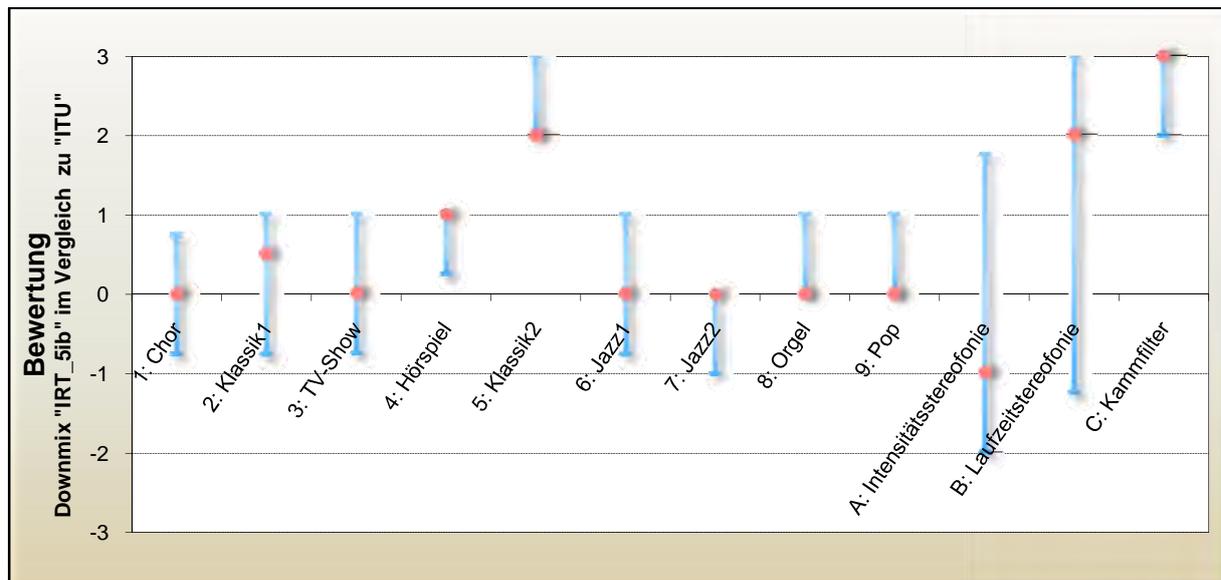


Abbildung 9.2-2 Downmix "IRT\_5ib" im Vergleich zu "ITU"

Im Vergleich zur ITU-Empfehlung zeigt die neue Variante des IRT-Downmixes („5ib“) deutlich bessere Ergebnisse beim zweiten Klassikbeispiel und den synthetisch generierten Beispielen B und C (Abbildung 9.2-2). Diese Ergebnisse veranschaulichen die effektive Wirksamkeit der Kammfilterkompensation. Selbst bei dem äußerst schwierigen Testmaterial mit der durch Laufzeitstereofonie erzeugten Phantomschallquelle (B) überzeugte das optimierte Verfahren. Auch in den Hörbeispielen „Klassik1“ und „Hörspiel“ konnte eine Verbesserung gegenüber der ITU-Empfehlung wahrgenommen werden.

Lediglich das Hörbeispiel mit der durch Pegeldifferenzen erzeugten Phantomschallquelle zeigt einen Medianwert mit negativem Vorzeichen. Der große Bereich zwischen den Quartilen deutet auf eine sehr unterschiedliche Bewertung hin. Möglicherweise ist die Ursache in der Bearbeitungsmethode der DFT-Koeffizienten in Fall 2 zu suchen. Die Variante „5ib“ verwendet dort noch nicht die in Kapitel 8 beschriebene Optimierung. Die Abbildungsstabilität der verwendeten alten Variante offenbart in Abbildung 8.1-2 (orangefarbener Graph) erhebliche Schwächen.

## 9.2.3 Vergleich IRT-Downmix „5j“ mit dem Downmix nach der ITU-Empfehlung

Der Vergleich der Downmixvariante „5j“ mit der ITU-Empfehlung zeigt ein ähnliches Ergebnis (Abbildung 9.2-3). Im Gegensatz zum vorangegangenen Vergleich zeigt die Variante „5j“ jedoch bei Hörbeispiel A ein positives Ergebnis. Wie bereits vermutet ist dies auf die in Variante „5j“ angewandte Optimierung der Abbildungsstabilität aus Kapitel 8 zurückzuführen.

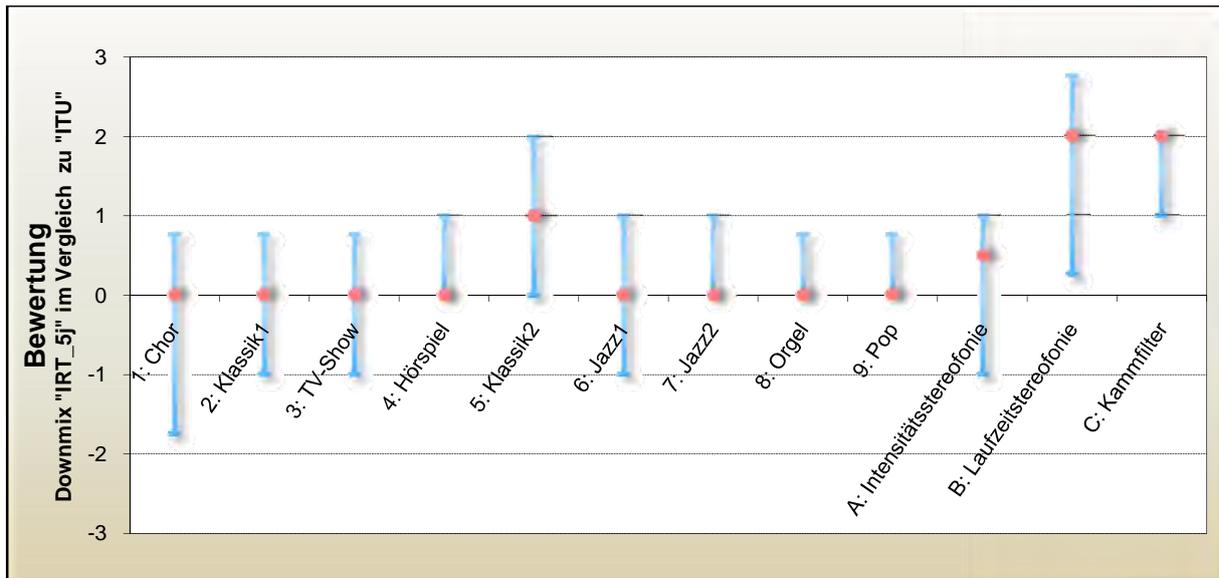


Abbildung 9.2-3 Downmix "IRT\_5j" im Vergleich zu "ITU"

#### 9.2.4 Vergleich IRT-Downmix „5ib“ mit dem IRT-Downmix „5j“

Der direkte Vergleich der Varianten „5ib“ und „5j“ offenbart eine tendenziell bessere Bewertung der Variante „5j“, mit Ausnahme des Hörbeispiels C (Abbildung 9.2-4). Diese Abwertung ist auf das wahrnehmbare Artefakt zurückzuführen, das bereits in Kapitel 7.3.5 erwähnt wurde.

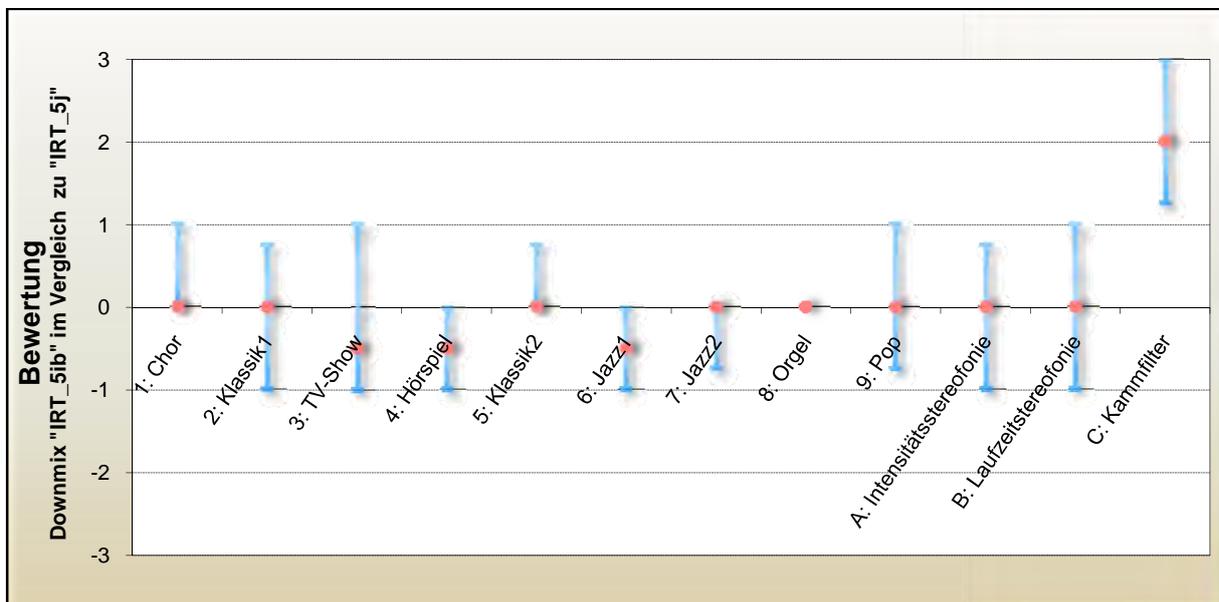


Abbildung 9.2-4 Downmix "IRT\_5ib" im Vergleich zu "IRT\_5j"

#### 9.2.5 Vergleich IRT-Downmix „5ib“ mit einer von Hand erstellten Stereoverision

Der letzte Paarvergleich stellt die Downmixvariante „5ib“ einem Handmix gegenüber. Dieser Vergleich ist nur bedingt möglich und kann lediglich ein sinnvolles Ergebnis hinsichtlich der technischen Qualität liefern. Künstlerische Aspekte können bei einem solchen Vergleich bei den Versuchsteilnehmern oft nur bedingt Anerkennung finden. Schließlich sucht der Proband in vielen

Fällen nach einem möglichst ähnlichen Abbild der Mehrkanalversion. Besondere Intensionen des Tonmeisters oder Toningenieurs für die Zweikanalversion können dabei sogar zu einer negativen Wertung führen.

Für die Evaluation und einen eventuellen zukünftigen Einsatz des automatischen Downmixverfahrens ist dieser Vergleich jedoch unerlässlich. Dabei liegt das Augenmerk auf der technischen Qualität, die mit der eines Handmixes vergleichbar sein sollte.

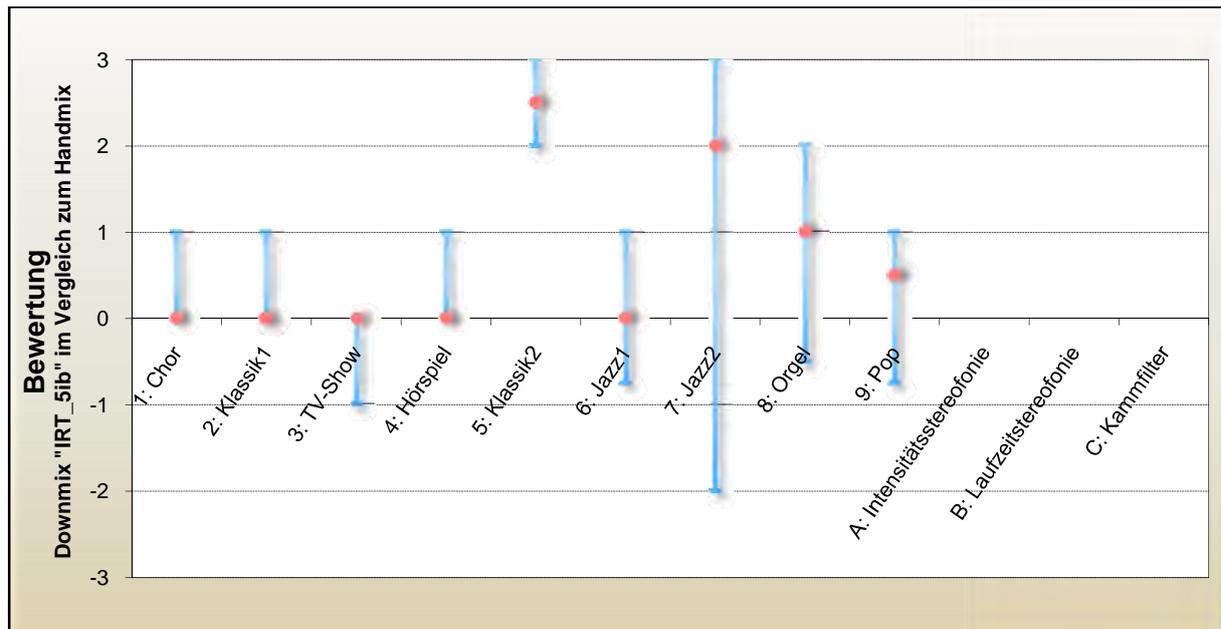


Abbildung 9.2-5 Downmix "IRT\_5ib" im Vergleich zum Handmix

Abbildung 9.2-5 zeigt ein hervorragendes Ergebnis für den optimierten IRT-Downmix. In keinem der Hörbeispiele wurde diese Downmixversion schlechter empfunden als die von Hand erstellte Version. In einigen Beispielen wird der IRT-Downmix sogar besser bewertet. Die große Spannweite der Quartile in Hörbeispiel 7 deutet auf ein gespaltenes Bewertungsverhalten bei den einzelnen Versuchsteilnehmern hin. Für die synthetischen Hörbeispiele entfällt dieser Vergleich.

### 9.2.6 Zusammenfassung der Ergebnisse

Mit jeweils einer Ausnahme bei einem synthetisch generierten Hörbeispiel konnten bei dem Hörversuch keine Schwächen der optimierten Algorithmen „5ib“ und „5j“ ausgemacht werden. Vor allem bei kritischem Ausgangsmaterial ist sogar eine deutliche positive Tendenz zu erkennen.

Durch die Kammfilterkompensation erzielen dort beide Lösungsansätze eine ausgeglichene Klangfarbe und Räumlichkeit. Die Variante „5j“ kann außerdem eine Verbesserung bei der Abbildungsstabilität vorweisen.

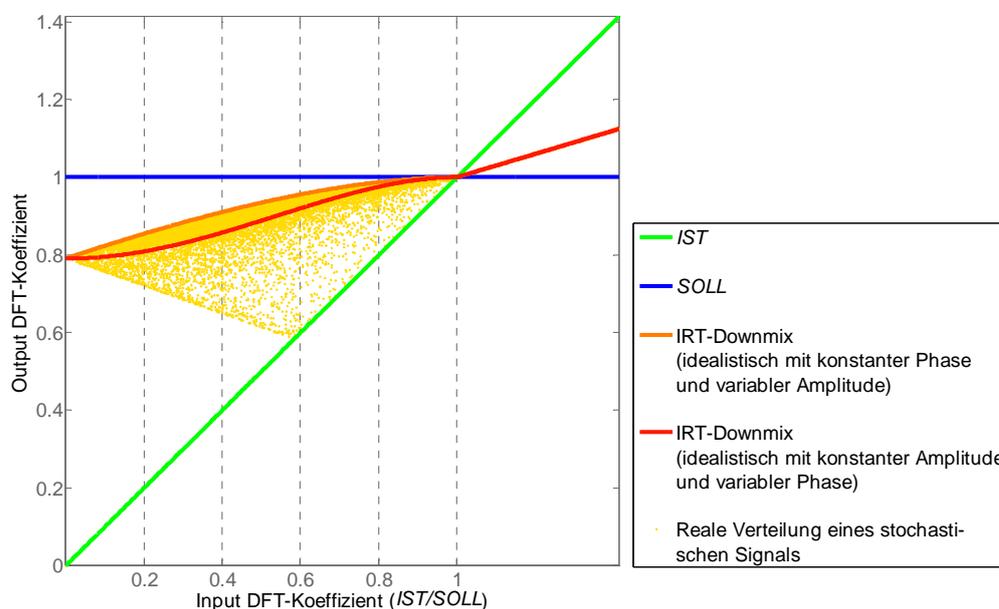
Im Vergleich zur Variante von Keil sind Verbesserungen bei der Artefaktbildung und der Klangfarbe klar erkennbar.

Generell kann festgestellt werden, dass ein Downmix immer auch von der Qualität der Mehrkanalmischung abhängt. Bei einer guten Mischung des Mehrkanaltons führt deshalb häufig auch schon ein einfacher ITU-Downmix zu akzeptablen Ergebnissen. Bei schwierigen Ausgangssituationen, wie sie in der Praxis oft nicht oder nur schwer zu vermeiden sind, liegt der optimierte IRT-Downmix klar im Vorteil.

### 9.3 Nachträgliche Anpassung des Algorithmus

Die Evaluation bestätigt im Wesentlichen die bereits favorisierte IRT-Downmix-Variante „5ib“. Das etwas schlechtere Ergebnis im Vergleich zur Variante „5j“ bei Hörbeispiel A ist auf die Optimierung aus Kapitel 8 zurückzuführen, die erst ab Version „5j“ in den Algorithmus implementiert ist.

Es ist daher naheliegend, diesen Vorteil bei der Abbildungsstabilität bereits bei Variante „5ib“ einzusetzen. Der Unterschied hinsichtlich der Abbildung liegt bei den beiden Varianten in der Bearbeitung der DFT-Koeffizienten in Fall 2. Die in Hörbeispiel C auftretenden Artefakte bei der Variante „5j“ sind auf die Bearbeitung in Fall 1 zurückzuführen. Aus diesem Grund kann die Bearbeitung in Fall 2 der Variante „5j“ mit der Bearbeitung in Fall 1 der Variante „5ib“ kombiniert werden. Mit dieser Veränderung werden die Vorteile der Varianten „5ib“ und „5j“ vereint und gleichzeitig die Schwachstellen beider Varianten verhindert. Das Ergebnis ist der IRT-Downmix „5r“, der abschließend alle im Rahmen dieser Arbeit erzielten Verbesserungen und Erkenntnisse beinhaltet. Abbildung 9.3-1 zeigt den Kennlinienverlauf der Variante „IRT\_5r“.



**Abbildung 9.3-1** Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix „5r“. Die Bearbeitung erfolgt mit einer Fallunterscheidung. Die Korrektur in Fall 1 ( $IST < SOLL$ ) erfolgt mit dem Korrekturfaktor  $x[p]$  aus (7.12), wobei  $w = 0,25$ . Für DFT-Koeffizienten, die mit der Bearbeitung eine Dämpfung erfahren, wird der zugehörige  $IST$ -Wert verwendet. Die Bearbeitung in Fall 2 erfolgt anhand der Optimierung in Kapitel 8.

## 10 Weitere Überlegungen hinsichtlich eines automatischen Downmixes

### 10.1 Lautheit und Dynamik von Stereo- und Mehrkanal-Produktionen

Der gemischte Sendebetrieb<sup>137</sup> moderner Rundfunkanstalten besteht zunehmend - in Bezug auf die Anzahl der Audiokanäle - aus unterschiedlichem Sendematerial. Neben Stereo- werden auch eine Reihe von Mehrkanal-Produktionen verbreitet. Dabei ist die Konsistenz der Stereo- und Mehrkanal-Mischungen differenziert zu betrachten. Neben der Anzahl der verwendeten Kanäle unterscheiden sie sich vor allem in ihrer Programmdynamik.<sup>138</sup>

#### 10.1.1 Zweikanal-Stereo

Bis heute werden Mono- und Stereoaufnahmen auf Grund der oft ungünstigen Abhörbedingungen hinsichtlich einer guten Verständlichkeit abgemischt. Das Ergebnis ist ein kleiner Dynamikumfang sowohl im Radio als auch im Fernsehen. Zusätzlich wird das Spektrum der einzelnen Hörereignisse eingegrenzt, um spektrale Verdeckungseffekte in der Mischung zu vermeiden.

#### 10.1.2 Mehrkanal

Im Gegensatz zu Zweikanal-Stereo geht man bei Mehrkanalproduktionen (5.0) von relativ guten Abhörbedingungen aus. Mit dem Kino als Ursprung und Vorbild, werden auch im Heimbereich mit geeigneten Lautsprechern und akustischen Maßnahmen in oft extra dafür vorgesehenen Räumlichkeiten spezielle Abhörsituationen geschaffen, die eine gute Wiedergabe von Mehrkanalton ermöglichen. Mehrkanalsignale mit einem größeren Dynamikumfang können damit auch zu Hause noch akzeptabel wiedergegeben werden.

Der 5.0-Mehrkanalton bietet im Gegensatz zur Abbildung über zwei Lautsprecher mehr Raum für die Anordnung und Staffelung der Hörereignisse. Die Ortung leiser oder zarter Hörereignisse ist deutlich leichter, wenn sie aus einer anderen Richtung als lautere oder breitbandigere Signale eintreffen. Bei der Wiedergabe über fünf Lautsprecher stehen für die Bildung von Phantomschallquellen fünf Paare benachbarter Lautsprecher zur Verfügung, so dass nicht jedes Hörereignis zwangsläufig über Signalanteile in allen Kanälen verfügen muss.

#### 10.1.3 Probleme beim automatischen Downmix

Ausgehend von diesen Überlegungen ist zu erwarten, dass eine durch einen automatischen Downmix erstellte Stereoersion im Vergleich zu einem speziell für Stereoanwendungen erstellten Handmix einen ungewohnt großen Dynamikumfang aufweist. Kombiniert man von Hand gemischte Stereoproduktionen mit Signalen, die durch einen automatischen Downmix entstanden sind, so können Schwankungen in der Lautheit und Dynamik auftreten.

---

<sup>137</sup> Im gemischten Sendebetrieb werden, je nach Ausgangsmaterial, sowohl zwei- als auch mehrkanalige Audiosignale verbreitet.

<sup>138</sup> vgl. Naumann 2008, S. 29

Der viel kleinere Abbildungsausschnitt bei der Stereowiedergabe kann dazu führen, dass sich verschiedene Signalanteile gegenseitig überlagern. Ein geflüsterter Dialog kann beispielsweise nach dem Downmix von einer breitbandigen Atmosphäre verdeckt werden. Abbildung 10.1-1 verdeutlicht dieses Problem grafisch.

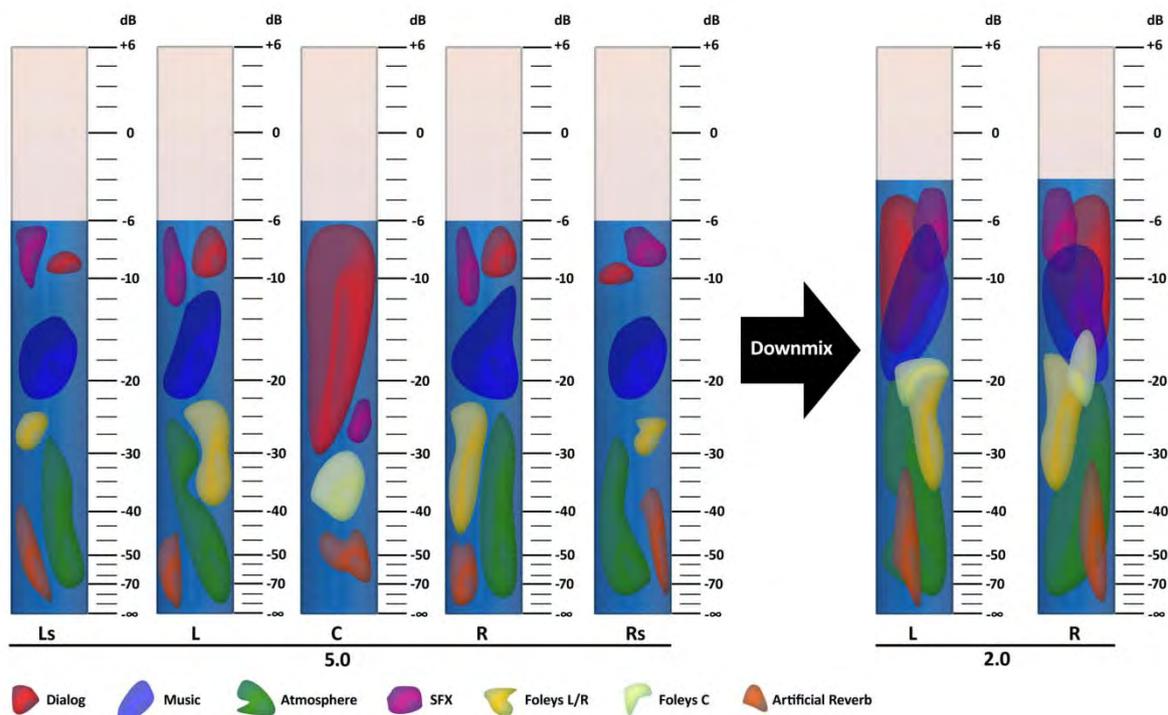


Abbildung 10.1-1 Verdeckungsprobleme beim automatischen Downmix von 5.0 auf 2.0 am Beispiel einer Filmmischung

Bei Popmusik finden sich in der Mehrkanalmischung oft zusätzliche Klangelemente, was zu einem attraktiveren Klangerlebnis führt. Nach einem automatischen Downmix besteht jedoch die Gefahr, dass eine ausreichende Abgrenzung dieser Elemente gegenüber essentiellen Signalanteilen nicht mehr gegeben ist. Die entstandene Stereoverision wirkt schnell überladen.

Dieses Kapitel entstand in Zusammenarbeit mit Silvio Naumann, der zeitgleich zu dieser Arbeit am Institut für Rundfunktechnik eine Diplomarbeit zum Thema „Entwicklung eines Verfahrens zur Generierung, Übertragung und Auswertung lautstärkerelevanter Kennwerte“ verfasst hat.

## 10.2 Verbreitung von Mehrkanal und Stereo

Diese Arbeit geht davon aus, dass ein Stereo-Downmix unter anderem die Abwärtskompatibilität einer Mehrkanaltonproduktion sichern soll. Aus dieser Annahme geht natürlich auch die Frage hervor, an welcher Stelle der Übertragungskette bis zum Rezipienten ein Downmix erfolgen soll und wie die Kompatibilität, also die Zugriffsmöglichkeit seitens des Rezipienten auf beide Tonversionen realisiert werden kann.

Eine Lösung wäre, acht diskrete Kanäle - sechs Kanäle für den 5.1 Mehrkanalton und zwei für den Stereoton - zu übertragen. Steht eine entsprechend große Bandbreite zur Verfügung, ist

diese Möglichkeit zu favorisieren. In vielen Fällen ist aber die zur Verfügung stehende Bandbreite knapp und reicht nicht zur simultanen und diskreten Übertragung beider Tonversionen.

Eine weitere Möglichkeit besteht darin, den Downmix empfängerseitig zu platzieren. In diesem Falle müsste nur die Mehrkanalversion übertragen werden, aus der dann bei Bedarf eine zweikanalige Version im Receiver erstellt wird. Ein „Upmix“ - also die Wiederherstellung des Mehrkanaltons aus einem Stereoton ist dabei nicht nötig. Es gibt jedoch auch einen Nachteil: die Erzeugung des Downmixes entzieht sich jeglicher Kontroll- und Anpassungsmöglichkeit. Parametereinstellungen für bestimmte Genres können nicht verändert werden. Mit dem Anspruch, einen sehr guten Stereo-Downmix zur Verfügung zu stellen, ist diese Variante erst einmal ungeeignet.<sup>139</sup>

Die dritte Möglichkeit liegt in der Übertragung des Stereotons neben einem zusätzlichen Steuersignal, das es einem empfängerseitigen Decoder erlaubt, die sechs diskreten Signale der Mehrkanalversion zu reproduzieren - also einen gesteuerten Upmix zu erstellen.

Einen solchen Ansatz verfolgt beispielsweise die MPEG-Surround-Codierung, die neben den zwei Kanälen des Downmixes sogenannte „Spatial Cues“ in den Datenstrom miteinschließt, die das Mehrkanalsignal in Bezug auf das vollwertig übertragene Downmix-Signal definieren.

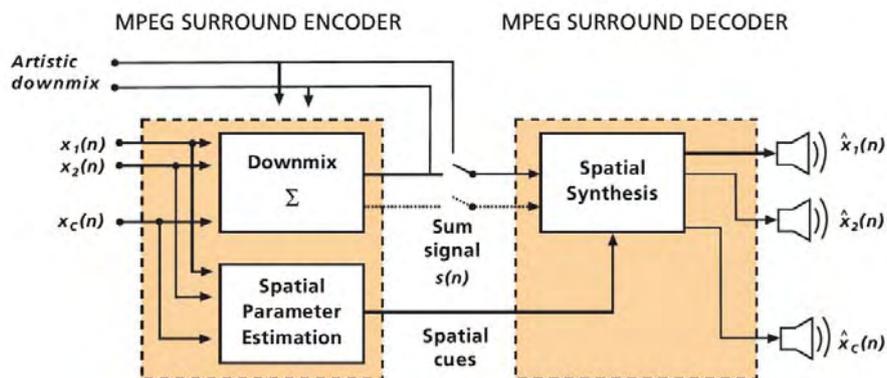


Abbildung 10.2-1 Prinzip der MPEG-Surround Codierung

Die „Spatial Cues“ werden von MPEG Surround, das gemeinschaftlich von Fraunhofer IIS/Agerer Systems und Coding Technologies/Philips entwickelt wurde, mit Hilfe von drei Eigenschaften der Psychoakustik ermittelt: die ITD (Interaural Time Difference), die Auskunft über den zeitlichen Versatz an beiden Ohren gibt, den ein Schallereignis hervorruft, die ILD (Interaural Level Difference), welche die aus einem Schallereignis resultierende Pegeldifferenz an beiden Ohren angibt, und die IC (Interaural Correlation), die Informationen über die Ähnlichkeit der beiden interauralen Signale liefert.<sup>140</sup>

<sup>139</sup> Unmöglich ist es nicht, den Downmixalgorithmus in Empfangsgeräte, wie zum Beispiel Receiver zu integrieren. Dabei wäre es denkbar die nötigen Parameter über mitgesendete Metadaten normiert zu steuern.

<sup>140</sup> vgl. Schick 2004, S. 5 ff.

### 10.2.1 Ein Radioprogramm im 5.1 Mehrkanal-Format

Abschließend soll das Szenario eines komplett in 5.1 Mehrkanalton ausgestrahlten Radioprogramms entworfen werden. Es ist anzunehmen, dass weder von allem zu sendenden Material eine Mehrkanalversion noch von allem eine Stereoverversion vorhanden ist. Im Sendebetrieb entstehen daher drei Möglichkeiten:

1. Von dem zu sendenden Material liegt sowohl eine Mehrkanal- als auch eine Stereoverversion vor.
2. Vom zu sendenden Material liegt nur eine Mehrkanalversion vor.
3. Vom zu sendenden Material liegt nur eine Stereoverversion vor.

Das in Abbildung 10.2-2 dargestellte Modell zeigt einen möglichen Lösungsansatz auf. Im ersten Fall würden nur die Steuerdaten zur empfängerseitigen Wiedergewinnung der sechs diskreten Kanäle generiert werden. Ein Down- oder Upmix wäre nicht erforderlich. Läge nur ein Mehrkanalton vor, so würde der Stereoton senderseitig durch einen Downmix erstellt. Wäre das Ausgangsmaterial nur in Stereo vorhanden, so würden die Steuerdaten den Bezug zu einem senderseitig manipulierbaren Upmix herstellen.

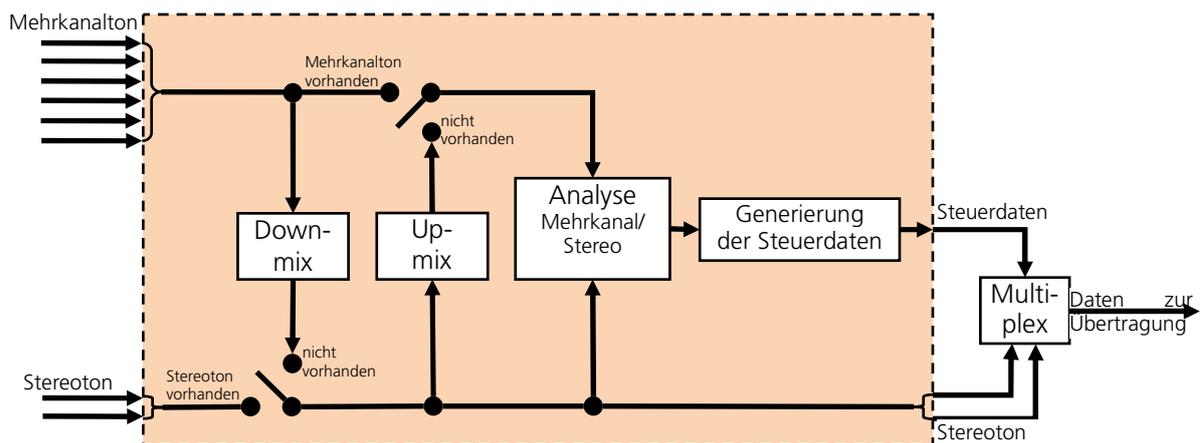


Abbildung 10.2-2 Modell zur Realisierung eines 5.1-Radioprogramms

## 11 Anwendungsmöglichkeiten

Vom Rundfunk über die Audioproduktion bis hin zur Downmix-Lösung für den Konsumenten bietet der IRT-Downmix, durch seine modulare Aufbauweise und die flexibel variierbaren Parameter, vielfältige Einsatzmöglichkeiten. Außerdem kann er - vom Downmixgedanken losgelöst - als Mischverfahren zweier Signale mit Kammfilterkompensation dienen.

### 11.1 Der IRT-Downmix beim Rundfunk

Seit Beginn des Downmix-Projekts am IRT ist es das Ziel, ein Verfahren zu finden, das dem Rundfunk erlaubt, bei hoher Qualität effizient eine Stereo-variante des Mehrkanaltons zu erstellen.<sup>141</sup>

Die vorliegende Lösung erlaubt die Anwendung des IRT-Downmixes sowohl bei der Produktion<sup>142</sup> als auch im Live-Betrieb vor der Ausstrahlung. Bei der zweiten Möglichkeit bietet es sich an, die in der Sendung verwendeten Audiosignale mit Steuerdaten zu verknüpfen, welche dann beim Downmix beispielsweise die Dämpfungsparameter für Center- und Surroundkanäle bestimmen. Gerade Laufplansendungen<sup>143</sup>, deren Inhalte oft erst kurzfristig zur Verfügung stehen, können auch im 5.1-Mehrkanalton verwirklicht werden. Der IRT-Downmix garantiert dabei die zeitgleiche Generierung des Stereotons.

Bei einem solchen Anwendungsfall ist es sinnvoll, den IRT-Downmix als Echtzeit-Hardwarelösung zu realisieren.

### 11.2 Der IRT-Downmix in der Audioproduktion

Sämtliche Audioproduktionen, bei denen eine Mischung im Mehrkanalton sinnvoll und eine Stereo-version notwendig ist, sind potentielle Anwendungsfelder für den IRT-Downmix. Mit einer zunehmenden Beliebtheit des mehrkanaligen Klangerlebnisses sehen sich Produzenten im Audio-bereich immer häufiger der Frage gegenüber, in welchem Tonformat produziert werden soll. Die Möglichkeit eines qualitativ hochwertigen Downmixes erleichtert die Entscheidung zugunsten des Mehrkanaltons. Die Anwendungsfelder erstrecken sich von der Filmtoneherstellung über die Musik- und Sprachproduktion bis hin zum Sounddesign für interaktive Inhalte wie beispielsweise Computerspiele.

Sowohl eine Lösung auf Hardwarebasis als auch eine Umsetzung als VST-Plug-in<sup>144</sup> können eingesetzt werden. Die Softwarelösung ermöglicht dabei eine flexible Integration in eine Reihe computerbasierter Produktionstechniken. Beide Varianten erlauben eine Echtzeit-Anwendung und können dem Tonmeister oder Toningenieur so bereits während der Produktion das Ergebnis des Downmixes zur Verfügung stellen. Dadurch wird durch ein Umschalten der Wiedergabe eine Kontrolle des Downmixes möglich.

---

<sup>141</sup> vgl. Stoll 2005, S. 50

<sup>142</sup> vgl. Kapitel 11.2 Der IRT-Downmix in der Audioproduktion

<sup>143</sup> Laufplansendungen sind moderierte Sendungen mit einem festen zeitlichen Ablauf von Beiträgen und Musiktiteln.

<sup>144</sup> Die Virtual Studio Technology (VST) wurde von der Firma Steinberg Media Technologies entwickelt und bietet eine Schnittstelle für Audio-Software, die den Dialog zwischen einem VST-Host (zum Beispiel Sequenzer-Programme) und einem entsprechenden Plug-in erlaubt.

### 11.3 Der IRT-Downmix beim Rezipienten

Bei Konsumergeräten macht die Funktionalität eines Downmixes ebenfalls Sinn. Wie bereits in Kapitel 10.2 angedacht, könnte der Algorithmus direkt in die Digitale Signalverarbeitung entsprechender Geräte wie Set-Top-Boxen, Verstärker oder Autoradios integriert werden. Die Ansteuerung könnte über in die Audiosignale eingebettete Metadaten erfolgen.

### 11.4 Der IRT-Downmix als generelles Mischverfahren zweier Signale

Auch wenn der Downmix den Anstoß zur Entwicklung eines Verfahrens zur Kammfilterkompensation gegeben hat, muss sein Einsatz nicht auf dieses Gebiet beschränkt bleiben. Sämtliche Mischvorgänge von Signalen, bei denen Interferenzen oder Kammfiltereffekte auftreten können, sind prädestiniert für eine Anwendung des Additionsverfahrens mit Kammfilterkompensation.

Gerade die Verwendung von Signalen in einer Mischung, die nicht vollständig dekorreliert sind, ruft diesen klangverfälschenden Effekt hervor. Durch den Einsatz der Kammfilterkompensation könnte in einem solchen Fall die Klangfarbe erhalten bleiben. Hier wäre eine Implementierung als Plug-in für digitale Mischpulte oder Sequenzer-Programme<sup>145</sup> denkbar.

---

<sup>145</sup> Ein Sequenzer-Programm ist eine Software zur Aufnahme, Wiedergabe und Bearbeitung von Audiosignalen. Vgl. Sequenzer (Musik) - Wikipedia 2008



## 12 Zusammenfassung und Ausblick

Die Auswertung mehrerer Untersuchungen von automatischen Downmixverfahren hat gezeigt, dass ein automatischer Downmix künstlerische Aspekte niemals ersetzen kann. Hinsichtlich ihrer technischen Qualität erreichen aktive Downmixverfahren jedoch ein Niveau, das durchaus mit einer von Hand erstellten Stereo-Mischung mithalten und ihr sogar überlegen sein kann.

Dabei sind die Herausforderungen, die sich einem Downmixverfahren stellen, keineswegs trivial. Laufzeitunterschiede kohärenter Anteile in den Signalen des Mehrkanaltons können bei der Matrizierung zu Kammfiltereffekten und somit zu einer starken Veränderung in der Klangfarbe führen.

Die unterschiedlichen Eigenschaften der elektrischen und akustischen Addition sowie der Addition von kohärenten und inkohärenten Signalanteilen führen zu Änderungen in der Lautstärkebalance und der Abbildung von Phantomschallquellen.

Ausgehend von der ITU-Empfehlung sieht der IRT-Downmix bei jeder Addition zweier Signale eine intelligente Bearbeitung im Spektralbereich vor. Diese Korrektur in kleinsten Frequenzbändern mit dem Ziel der Kammfilterkompensation gewährleistet nicht nur die Erhaltung der Klangfarbe des Audiosignals, sondern auch die Lautstärkebalance und die Abbildungstabilität im Frontbereich.

Die Parameter zur Dämpfung des Centersignals und der Surroundsignale bleiben variabel, so dass sie flexibel an unterschiedliche Genres angepasst werden können.

Die im Rahmen dieser Arbeit entstandene Variante des IRT-Downmixes kann mit einer noch ausgeglicheneren Klangfarbe und einem angenehm präsenten Höreindruck neue Akzente setzen. Fehler konnten beseitigt und die Bildung von Artefakten auf ein nicht wahrnehmbares Minimum reduziert werden.

Die Ergebnisse des Hörversuchs zeigen, dass der IRT-Downmix im Vergleich zu einer von Hand erstellten Stereoersion beim Hörer keine wahrnehmbaren Einbußen hervorruft. Auch wenn die Unterschiede bei manchen Hörbeispielen gegenüber den Konkurrenten gering sind, zeigen sich vor allem bei kritischem Ausgangsmaterial die Qualitäten des IRT-Downmixes deutlich. Damit hat das Verfahren einen Zustand erreicht, mit dem es für den praktischen Einsatz gerüstet ist.

Sowohl Anwendungen bei Hörfunk und Fernsehen als auch in der Audioproduktion und unter Live-Bedingungen bietet sich der Einsatz dieses neuen Verfahrens an. Mit der wachsenden Verbreitung des Mehrkanaltons kann der IRT-Downmix die Kompatibilität zu Stereo mit guter Qualität sichern. Darüber hinaus kann der Algorithmus bei sämtlichen Matrizier- oder Signaladditionsvorgängen eingesetzt werden und die Entstehung von Kammfiltern wirksam unterbinden.

Weitere Optimierungsmöglichkeiten der vorliegenden Lösungsansätze könnten im Bereich der Phasenprädiktion und der Abbildung von Phantomschallquellen liegen. Beides wäre jedoch mit einem nicht unerheblichen Mehraufwand an Rechenleistung verbunden. Außerdem darf bezweifelt werden, ob solch eine weitere Optimierung zu einem noch besseren, subjektiven Höreindruck führt.

Neue Ansätze sind in der Trennung der im Signal enthaltenen Hörereignisse zu suchen. Liegen die einzelnen Hörereignisse als diskrete Signale vor, können sie gezielt für die Zweikanalwieder-

gabe wieder zusammengesetzt werden. Besonders eine sinnvolle Anordnung der Phantomschallquellen, die Verhinderung von Verdeckungseffekten und eine gezielte dynamische Bearbeitung einzelner Anteile könnten auf diese Weise ermöglicht werden.

Eine umfangreiche Evaluation, wie sie im Rahmen dieser Arbeit leider nicht möglich war, ist in jedem Fall zu empfehlen. Neben einer detaillierteren Verifikation des IRT-Downmixes, könnte sie vor allem die Akzeptanz eines solchen Verfahrens im Rundfunk- und Audiobereich erhöhen und die erzielten Erfolge publik machen.

## 13 Anhang

13	Anhang.....	102
13.1	Quelltext .....	103
13.1.1	Quelltext: IRT-Downmix, Version „IRT_5keil“ .....	103
13.1.2	Quelltext: IRT-Downmix, Version „IRT_5ib“ .....	107
13.1.3	Quelltext: IRT-Downmix, Version „IRT_5j“ .....	111
13.1.4	Quelltext: IRT-Downmix, Version „IRT_5r“ .....	115
13.1.5	Direkter Vergleich der Algorithmen .....	119
13.1.6	Quelltext: Funktion zur Darstellung der Bearbeitung der DFT-Koeffizienten.....	121
13.1.7	Quelltext: Ausgabe einer Fallanalyse zweier DFT-Koeffizienten .....	123
13.2	Berechnung der Pegel einer Phantomschallquelle abhängig vom Winkel.....	125
13.3	Blockschaltbild des IRT-Downmix.....	127
	Analyse und Korrektur 1 .....	128
	Analyse und Korrektur 2 .....	128
	Analyse und Korrektur 3 .....	129
	Analyse und Korrektur 4 .....	129
13.4	Schriftliche Einführung für den Hörversuch .....	130

### 13.1.1 Quelltext: IRT-Downmix, Version „IRT\_5keil“

Der folgende Quelltext beinhaltet qualitativ die Downmixversion von Martin Keil in der modularen Form von Bernfried Runow.

```
IRT_5keil.m
```

```

1  %% Downmixverfahren IRT (Version IRT_5keil)
2  function IRT_5keil(verzeichnis,bezeichnung)
3  % IRT-Downmix: Matriziert ein 5.0 Mehrkanalsignal auf ein Stereosignal. Der Additions-
4  % prozess erfolgt mit einer intelligenten Kammfilterkompensation.
5  % function IRT_5keil(verzeichnis,bezeichnung)
6  % verzeichnis :Verzeichnisname als String, in dem die Mono-wav-files des
7  % Mehrkanalsignals liegen.
8  % bezeichnung :Bezeichnungskopf der Mono-wav-files des Mehrkanalsignals,
9  % der bei allen fünf Dateien identisch ist. Ihm folgt die
10 % Unterscheidung der einzelnen Kanäle mit L,R,C,SL,SR.
11 %IRT_5keil.m * Martin Keil, (Bernfried Runow) * 03/07/2008
12
13
14
15 %% Mischung L-C
16 % Einlesen der Mehrkanal-Audiosignale im .wav Format
17 pfad = strcat(verzeichnis,bezeichnung);
18 [l1,Fs,bits] = wavread(strcat(pfad,'L.wav'));
19 [c1,Fs,bits] = wavread(strcat(pfad,'C.wav'));
20
21
22 % An den Beginn und ans Ende der eingelesenen Signaldateien wird ein Nullvektor mit
23 % einer Länge von 1024 Samples gefügt. Ebenfalls an das Ende angefügt wird eine Reihe
24 % von Nullen "auffuell", so dass der Vektor genau durch 1024 teilbar wird.
25 auffuell = (fix(length(l1)/1024)+1)*1024-length(l1);
26
27 l = 0;
28 c = 0;
29
30 l(1025:(length(l1)+1024)) = l1;
31 c(1025:(length(c1)+1024)) = c1;
32
33 clear l1 c1;
34
35 l((length(l)+1):(length(l)+auffuell+1024)) = 0;
36 c((length(c)+1):(length(c)+auffuell+1024)) = 0;
37
38
39 % An dieser Stelle wird der Vektor zur weiteren Verarbeitung vom Zeilen ins
40 % Spaltenformat gedreht
41 l = l';
42 c = c';
43
44
45 % Zunächst wird das Centersignal um -3 dB im Pegel abgesenkt
46 c = c * (1/sqrt(2));
47
48
49 % Festlegung, wie oft ein halbes Fenster in den Audiostream passt
50 fg = fix(length(l)/1024);
51
52
53 % Mischung wird durchgeführt
54 erg = mix(l,c,1,Fs,fg);
55
56
57 % Das gemischte Signal wird als wav-file gespeichert
58 wavwrite(erg,Fs,bits,'c:\temp\lc');
59
60 clear l;
61
62
63
64 %% Mischung C-R
65
66 [r1,Fs,bits] = wavread(strcat(pfad,'R.wav'));
67
68 r=0;
69 r(1025:(length(r1)+1024)) = r1;
70 clear r1;
71 r((length(r)+1):(length(r)+auffuell+1024)) = 0;

```

```

72 r = r';
73
74 erg = mix(r,c,2,Fs,fg);
75 wavwrite(erg,Fs,bits,'c:\temp\rc');
76
77 clear r c;
78
79
80
81 %% Mischung LC-LS
82
83 [lc,Fs,bits] = wavread('C:\temp\lc.wav');
84 [ls1,Fs,bits] = wavread(strcat(pfad,'SL.wav'));
85
86 ls = 0;
87 ls(1025:(length(ls1)+1024)) = ls1;
88 clear ls1;
89 ls((length(ls)+1):(length(ls)+auffuell+1024)) = 0;
90 ls = ls';
91 ls = ls * (1/sqrt(2));
92
93 erg = mix(lc,ls,3,Fs,fg);
94 erg = erg(1025:length(erg)-(3072+auffuell));
95 wavwrite(erg,Fs,bits,'c:\output\links');
96
97 clear ls lc;
98
99
100
101 %% Mischung RC-RS
102
103 [rc,Fs,bits] = wavread('C:\temp\rc.wav');
104 [rs1,Fs,bits] = wavread(strcat(pfad,'SR.wav'));
105
106 rs = 0;
107 rs(1025:(length(rs1)+1024)) = rs1;
108 clear rs1;
109 rs((length(rs)+1):(length(rs)+auffuell+1024)) = 0;
110 rs = rs';
111 rs = rs * (1/sqrt(2));
112
113 erg = mix(rc,rs,4,Fs,fg);
114 erg = erg(1025:length(erg)-(3072+auffuell));
115 wavwrite(erg,Fs,bits,'c:\output\rechts');
116
117 clear rs rc;
118
119
120 %% Stereo-Output
121
122 % es wird überprüft, ob das Verzeichnis 'encodiert' vorhanden ist
123 speicher_pfad = strcat(verzeichnis,'encodiert\');
124 if (isdir(speicher_pfad) == 0)
125     mkdir(speicher_pfad);
126 end
127
128 % damit's schöner aussieht: Zahlen mit nur einer Ziffer wird eine '0' vorangestellt.
129 Uhr = clock;
130
131 monat = int2str(Uhr(2));
132 if (length(monat)==1)
133     monat = strcat('0',monat);
134 end
135
136 tag = int2str(Uhr(3));
137 if (length(tag)==1)
138     tag = strcat('0',tag);
139 end
140
141 speicher_datei
142 strcat(speicher_pfad,bezeichnung,int2str(Uhr(1)),monat,tag,'_IRT#5keil');
143 wavwrite([wavread('c:\output\links') erg],Fs,bits,speicher_datei);
144
145 clear erg monat tag Uhr speicher_pfad speicher_datei
146
147
148
149

```

```
150 %% Funktion zum Mischen zweier Signale
151
152 function erg = mix(signal_1,signal_2,mixNr,Fs,fg)
153 % zwei Signale werden mittels des Algorithmus addiert
154
155 % Erzeugung eines Hannungfensters der Länge 2048 Samples
156 fe = hann(2048);
157
158
159 % Variablen für die Position von Beginn und Ende des Fensters im
160 % Audiosignal
161 f1 = 1;
162 f2 = 2048;
163
164
165 % Startvariable Fensterung
166 f = 0;
167
168
169 % Festlegung der Ergebnisdatei für die erste Mischung nach ihrer
170 % Bearbeitung
171 erg(1:3072) = 0;
172 erg          = erg';
173
174
175 % Start der Schleife für die Fensterung
176 while f < fg-1
177
178
179 % Fensterung mit gleichzeitiger Hüllkurvenbewertung durch das
180 % Hannungfenster
181 f_signal_1 = signal_1(f1:f2).*fe;
182 f_signal_2 = signal_2(f1:f2).*fe;
183
184
185 % Zero-Padding der gefensternten Signale
186 f_signal_1(2049:4096) = 0;
187 f_signal_2(2049:4096) = 0;
188
189
190 % Durchführung der FFT
191 spektrum_1 = fft(f_signal_1);
192 spektrum_2 = fft(f_signal_2);
193
194
195 % Anpassen der Spektren
196 spektrum_1 = spektrum_ap(spektrum_1);
197 spektrum_2 = spektrum_ap(spektrum_2);
198
199
200 % Bildung des Summensignals nach der Anpassung
201 spektrum_S = spektrum_1+spektrum_2;
202
203
204 % FFT-Koeffizienten werden bearbeitet
205 spektrum_S = koef_bearbeiten(spektrum_1, spektrum_2, spektrum_S, mixNr, f);
206
207
208 % Rücktransformation des bearbeiteten Summensignals in den Zeitbereich
209 signal_S   = ifft(spektrum_S);
210
211
212 % Zusammenfügen der Fenster:
213 % Aus den gefensternten Blöcken wird hier wieder ein Audiosignal zusammengesetzt.
214 position          = length(erg)-3071;
215 erg(position:(length(erg))) = erg(position:length(erg))+signal_S(1:3072);
216 erg(length(erg)+1:length(erg)+1024) = signal_S(3073:4096);
217
218
219 % Position für nächstes Fenster wird festgelegt
220 f1 = f1+1024;
221 f2 = f2+1024;
222
223 f = f+1;
224
225 end
226
227
```

```
228 %% Funktion zur Anpassung der Spektren
229
230 function spektrum = spektrum_ap(spektrum)
231 % Anpassen des Spektrums für spektrum; Nullsetzen des redundanten Anteils
232 n=2; % Gleichanteil wird nicht bearbeitet
233
234 % Die vordere Hälfte des Spektrums wird bis auf den ersten Wert (Gleichanteil) in der
235 % Amplitude verdoppelt
236 laenge = length(spektrum);
237
238 while (n<laenge/2+1)
239     spektrum(n) = spektrum(n)*2;
240     n=n+1;
241 end
242
243 % Die hintere Hälfte des Spektrums wird auf Null gesetzt
244 while (n<laenge+1)
245     spektrum(n) = spektrum(n)*0;
246     n=n+1;
247 end
248
249
250
251 %% Funktion zur Bearbeitung der FFT-Koeffizienten
252
253 function spektrumS = koef_bearbeiten(spektrum1, spektrum2, spektrumS, mixNr, f)
254 % Hier wird eine Fallunterscheidung vorgenommen und je nach Situation eines der drei
255 % Signale an dieser Stelle des Spektrums angepasst
256
257 n=2;
258 while (n < (length(spektrumS)/2+1))
259
260 % Einlesen der Amplituden an jeder Stelle der FFT der drei Signale
261     amplitudel = abs(spektrum1(n));
262     amplitudel2 = abs(spektrum2(n));
263     ist = abs(spektrumS(n));
264     soll = sqrt(amplitudel^2+amplitudel2^2);
265
266 % ...sowie der Winkel
267     winkel_ist = angle(spektrumS(n));
268     winkel_1 = angle(spektrum1(n));
269
270 % Fall 1
271     if ist < 0.1*soll
272         amplitudel_neu = amplitudel*1.9;
273         spektrum1(n) = amplitudel_neu*exp(winkel_1*i);
274         spektrumS(n) = spektrum1(n)+spektrum2(n);
275
276 % Fall 2
277     elseif ist < 0.6*soll && ist >= 0.1*soll
278         faktor = (soll/ist)*0.05+1.4;
279         amplitudel_neu = amplitudel*faktor;
280         spektrum1(n) = amplitudel_neu*exp(winkel_1*i);
281         spektrumS(n) = spektrum1(n)+spektrum2(n);
282
283 % Fall 4
284     elseif ist > soll
285         amplitudes_neu = soll+(ist-soll)*0.1;
286         spektrumS(n) = amplitudes_neu*exp(winkel_ist*i);
287
288 % Tritt keiner der oberen Fälle ein, so erfolgt keine Anpassung (Fall 3)
289     else
290     end
291     n = n+1;
292 end
```

## 13.1.2 Quelltext: IRT-Downmix, Version „IRT\_5ib“

```

                                                                    IRT_5ib.m
1  %% Downmixverfahren IRT (Version IRT_5ib)
2  function IRT_5ib(verzeichnis,bezeichnung)
3  % IRT-Downmix: Matriziert ein 5.0 Mehrkanalsignal auf ein Stereosignal. Der Additions-
4  % prozess erfolgt mit einer intelligenten Kammfilterkompensation.
5  % function IRT_5ib(verzeichnis,bezeichnung)
6  % verzeichnis :Verzeichnisname als String, in dem die Mono-wav-files des
7  % Mehrkanalsignals liegen.
8  % bezeichnung :Bezeichnungskopf der Mono-wav-files des Mehrkanalsignals,
9  %               der bei allen fünf Dateien identisch ist. Ihm folgt die
10 %               Unterscheidung der einzelnen Kanäle mit L,R,C,SL,SR.
11 %IRT_5ib.m * Bernfried Runow (Martin Keil) * 03/07/2008
12
13
14
15 %% Mischung L-C
16 % Einlesen der Mehrkanal-Audiosignale im .wav Format
17 pfad = strcat(verzeichnis,bezeichnung);
18 [l1,Fs,bits] = wavread(strcat(pfad,'L.wav'));
19 [c1,Fs,bits] = wavread(strcat(pfad,'C.wav'));
20
21
22 % An den Beginn und ans Ende der eingelesenen Signaldateien wird ein Nullvektor mit
23 % einer Länge von 1024 Samples gefügt. Ebenfalls an das Ende angefügt wird eine Reihe
24 % von Nullen "auffuell", so dass der Vektor genau durch 1024 teilbar wird.
25 auffuell = (fix(length(l1)/1024)+1)*1024-length(l1);
26
27 l = 0;
28 c = 0;
29
30 l(1025:(length(l1)+1024)) = l1;
31 c(1025:(length(c1)+1024)) = c1;
32
33 clear l1 c1;
34
35 l((length(l)+1):(length(l)+auffuell+1024)) = 0;
36 c((length(c)+1):(length(c)+auffuell+1024)) = 0;
37
38
39 % An dieser Stelle wird der Vektor zur weiteren Verarbeitung vom Zeilen ins
40 % Spaltenformat gedreht
41 l = l';
42 c = c';
43
44
45 % Zunächst wird das Centersignal um -3 dB im Pegel abgesenkt
46 c = c * (1/sqrt(2));
47
48
49 % Festlegung, wie oft ein halbes Fenster in den Audiostream passt
50 fg = fix(length(l)/1024);
51
52
53 % Mischung wird durchgeführt
54 erg = mix(l,c,1,Fs,fg);
55
56
57 % Das gemischte Signal wird als wav-file gespeichert
58 wavwrite(erg,Fs,bits,'c:\temp\lc');
59
60 clear l;
61
62
63
64 %% Mischung C-R
65
66 [r1,Fs,bits] = wavread(strcat(pfad,'R.wav'));
67
68 r=0;
69 r(1025:(length(r1)+1024)) = r1;
70 clear r1;
71 r((length(r)+1):(length(r)+auffuell+1024)) = 0;
72 r = r';

```

```

73
74   erg = mix(r,c,2,Fs,fg);
75   wavwrite(erg,Fs,bits,'c:\temp\rc');
76
77   clear r c;
78
79
80
81   %% Mischung LC-LS
82
83   [lc,Fs,bits] = wavread('C:\temp\lc.wav');
84   [ls1,Fs,bits] = wavread(strcat(pfad,'SL.wav'));
85
86   ls = 0;
87   ls(1025:(length(ls1)+1024)) = ls1;
88   clear ls1;
89   ls((length(ls)+1):(length(ls)+auffuell+1024)) = 0;
90   ls = ls';
91   ls = ls * (1/sqrt(2));
92
93   erg = mix(lc,ls,3,Fs,fg);
94   erg = erg(1025:length(erg)-(1024+auffuell));
95   wavwrite(erg,Fs,bits,'c:\output\links');
96
97   clear ls lc;
98
99
100
101   %% Mischung RC-RS
102
103   [rc,Fs,bits] = wavread('C:\temp\rc.wav');
104   [rs1,Fs,bits] = wavread(strcat(pfad,'SR.wav'));
105
106   rs = 0;
107   rs(1025:(length(rs1)+1024)) = rs1;
108   clear rs1;
109   rs((length(rs)+1):(length(rs)+auffuell+1024)) = 0;
110   rs = rs';
111   rs = rs * (1/sqrt(2));
112
113   erg = mix(rc,rs,4,Fs,fg);
114   erg = erg(1025:length(erg)-(1024+auffuell));
115   wavwrite(erg,Fs,bits,'c:\output\rechts');
116
117   clear rs rc;
118
119
120   %% Stereo-Output
121
122   % es wird überprüft, ob das Verzeichnis 'encodiert' vorhanden ist
123   speicher_pfad = strcat(verzeichnis,'encodiert\');
124   if (isdir(speicher_pfad) == 0)
125       mkdir(speicher_pfad);
126   end
127
128   % damit's schöner aussieht: Zahlen mit nur einer Ziffer wird eine '0' vorangestellt.
129   Uhr = clock;
130
131   monat = int2str(Uhr(2));
132   if (length(monat)==1)
133       monat = strcat('0',monat);
134   end
135
136   tag = int2str(Uhr(3));
137   if (length(tag)==1)
138       tag = strcat('0',tag);
139   end
140
141   speicher_datei
142   strcat(speicher_pfad,bezeichnung,int2str(Uhr(1)),monat,tag,'_IRT#5ib');
143   wavwrite([wavread('c:\output\links') erg],Fs,bits,speicher_datei);
144
145   clear erg monat tag Uhr speicher_pfad speicher_datei
146
147
148
149
150

```

```

151 %% Funktion zum Mischen zweier Signale
152
153 function erg = mix(signal_1,signal_2,mixNr,Fs,fg)
154 % zwei Signale werden mittels des Algorithmus addiert
155
156 % Erzeugung eines Hanningfensters der Länge 2048 Samples
157 fe = hann(2048);
158
159 % Variablen für die Position von Beginn und Ende des Fensters im Audiosignal
160 f1 = 1;
161 f2 = 2048;
162
163 % Startvariable Fensterung
164 f = 0;
165
166 % Festlegung des Ergebnisarrays für die Mischung nach ihrer Bearbeitung
167 erg(1:3072) = 0;
168 erg      = erg';
169
170 % Instanzieren der Fenstervariablen f_signal_1 und f_signal_2
171 f_signal_1 = [0 0];
172 f_signal_2 = [0 0];
173 f_signal_1 = f_signal_1';
174 f_signal_2 = f_signal_2';
175
176
177 % Start der Schleife für die Fensterung
178 while f < fg-1
179
180 % Zero-Padding, Fensterung mit gleichzeitiger Hüllkurvenbewertung durch das
181 % Hanningfenster
182 f_signal_1(1:1024) = 0;
183 f_signal_2(1:1024) = 0;
184 f_signal_1(1025:3072) = signal_1(f1:f2).*fe;
185 f_signal_2(1025:3072) = signal_2(f1:f2).*fe;
186 f_signal_1(3073:4096) = 0;
187 f_signal_2(3073:4096) = 0;
188
189
190 % Durchführung der FFT
191 spektrum_1 = fft(f_signal_1);
192 spektrum_2 = fft(f_signal_2);
193
194
195 % Anpassen der Spektren
196 spektrum_1 = spektrum_ap(spektrum_1);
197 spektrum_2 = spektrum_ap(spektrum_2);
198
199
200 % Bildung des Summensignals nach der Anpassung
201 spektrum_S = spektrum_1+spektrum_2;
202
203
204 % FFT-Koeffizienten werden bearbeitet
205 spektrum_S = koef_bearbeiten(spektrum_1, spektrum_2, spektrum_S, mixNr, f);
206
207
208 % Rücktransformation des bearbeiteten Summensignals in den Zeitbereich
209 signal_S = real(ifft(spektrum_S));
210
211
212 % Zusammenfügen der Fenster:
213 % Aus den gefensterten Blöcken wird hier wieder ein Audiosignal zusammengesetzt.
214 position = length(erg)-3071;
215 erg(position:(length(erg))) = erg(position:length(erg))+signal_S(1:3072);
216 erg(length(erg)+1:length(erg)+1024) = signal_S(3073:4096);
217
218
219 % Position für nächstes Fenster wird festgelegt
220 f1 = f1+1024;
221 f2 = f2+1024;
222
223 f = f+1;
224
225 end
226
227 erg = erg(1025:length(erg)-1024);
228

```

```
229 %% Funktion zur Anpassung der Spektren
230
231 function spektrum = spektrum_ap(spektrum)
232 % Anpassen des Spektrums für spektrum; Nullsetzen des redundanten Anteils
233 laenge = length(spektrum);
234
235 % Die vordere Hälfte des Spektrums wird bis auf den ersten Wert (Gleichanteil) in der
236 % Amplitude verdoppelt
237 spektrum(2:laenge/2) = spektrum(2:laenge/2)*2;
238
239 %Die zweite Hälfte wird gleich null gesetzt.
240 spektrum(laenge/2+1:laenge) = 0;
241
242
243
244 %% Funktion zur Bearbeitung der FFT-Koeffizienten
245
246 function spektrumS = koef_bearbeiten(spektrum1, spektrum2, spektrumS, mixNr, f)
247 % Hier wird eine Fallunterscheidung vorgenommen und je nach Situation eines der drei
248 % Signale an dieser Stelle des Spektrums angepasst
249
250 n=1;
251 while (n <= (length(spektrumS)/2))
252
253 % Einlesen der Amplituden an jeder Stelle der FFT der drei Signale
254     amplitudel = abs(spektrum1(n));
255     amplituden2 = abs(spektrum2(n));
256     ist = abs(spektrumS(n));
257     soll = sqrt(amplitudel^2+amplituden2^2);
258
259 % ...sowie der Winkel
260     winkel_ist = angle(spektrumS(n));
261     winkel_1 = angle(spektrum1(n));
262
263 % Fall 1
264     if (ist >= soll)
265         amplitudenS_neu = soll+(ist-soll)*0.1;
266         spektrumS(n) = amplitudenS_neu*exp(winkel_ist*i);
267
268 % Fall 2
269     elseif (ist < soll)
270         p = real(spektrum1(n))*real(spektrum2(n))+imag(spektrum1(n))*imag(spektrum2(n));
271         faktor = (-p+sqrt(p^2/4+amplitudel^4))/amplitudel^2;
272         amplituden1_neu = amplitudel*faktor;
273         spektrum1(n) = amplituden1_neu*exp(winkel_1*i);
274         spektrum_neu = spektrum1(n)+spektrum2(n);
275         if (abs(spektrum_neu)<=ist)
276             % Ist-Zustand bleibt erhalten
277         else
278             spektrumS(n)=spektrum_neu;
279         end
280     end
281     n = n+1;
282 end
```

## 13.1.3 Quelltext: IRT-Downmix, Version „IRT\_5j“

IRT\_5j.m

```

1  %% Downmixverfahren IRT (Version IRT_5j)
2  function IRT_5j(verzeichnis,bezeichnung)
3  % IRT-Downmix: Matriziert ein 5.0 Mehrkanalsignal auf ein Stereosignal. Der Additions-
4  % prozess erfolgt mit einer intelligenten Kammfilterkompensation.
5  % function IRT_5j(verzeichnis,bezeichnung)
6  % verzeichnis :Verzeichnisname als String, in dem die Mono-wav-files des
7  % Mehrkanalsignals liegen.
8  % bezeichnung :Bezeichnungskopf der Mono-wav-files des Mehrkanalsignals,
9  %               der bei allen fünf Dateien identisch ist. Ihm folgt die
10 %               Unterscheidung der einzelnen Kanäle mit L,R,C,SL,SR.
11 %IRT_5j.m * Bernfried Runow (Martin Keil) * 03/07/2008
12
13
14
15 %% Mischung L-C
16 % Einlesen der Mehrkanal-Audiosignale im .wav Format
17 pfad = strcat(verzeichnis,bezeichnung);
18 [l1,Fs,bits] = wavread(strcat(pfad,'L.wav'));
19 [c1,Fs,bits] = wavread(strcat(pfad,'C.wav'));
20
21
22 % An den Beginn und ans Ende der eingelesenen Signaldateien wird ein Nullvektor mit
23 % einer Länge von 1024 Samples gefügt. Ebenfalls an das Ende angefügt wird eine Reihe
24 % von Nullen "auffuell", so dass der Vektor genau durch 1024 teilbar wird.
25 auffuell = (fix(length(l1)/1024)+1)*1024-length(l1);
26
27 l = 0;
28 c = 0;
29
30 l(1025:(length(l1)+1024)) = l1;
31 c(1025:(length(c1)+1024)) = c1;
32
33 clear l1 c1;
34
35 l((length(l)+1):(length(l)+auffuell+1024)) = 0;
36 c((length(c)+1):(length(c)+auffuell+1024)) = 0;
37
38
39 % An dieser Stelle wird der Vektor zur weiteren Verarbeitung vom Zeilen ins
40 % Spaltenformat gedreht
41 l = l';
42 c = c';
43
44
45 % Zunächst wird das Centersignal um -3 dB im Pegel abgesenkt
46 c = c * (1/sqrt(2));
47
48
49 % Festlegung, wie oft ein halbes Fenster in den Audiostream passt
50 fg = fix(length(l)/1024);
51
52
53 % Mischung wird durchgeführt
54 erg = mix(l,c,1,Fs,fg);
55
56
57 % Das gemischte Signal wird als wav-file gespeichert
58 wavwrite(erg,Fs,bits,'c:\temp\lc');
59
60 clear l;
61
62
63
64 %% Mischung C-R
65
66 [r1,Fs,bits] = wavread(strcat(pfad,'R.wav'));
67
68 r=0;
69 r(1025:(length(r1)+1024)) = r1;
70 clear r1;
71 r((length(r)+1):(length(r)+auffuell+1024)) = 0;
72 r = r';

```

```
73
74 erg = mix(r,c,2,Fs,fg);
75 wavwrite(erg,Fs,bits,'c:\temp\rc');
76
77 clear r c;
78
79
80
81 %% Mischung LC-LS
82
83 [lc,Fs,bits] = wavread('C:\temp\lc.wav');
84 [ls1,Fs,bits] = wavread(strcat(pfad,'SL.wav'));
85
86 ls = 0;
87 ls(1025:(length(ls1)+1024)) = ls1;
88 clear ls1;
89 ls((length(ls)+1):(length(ls)+auffuell+1024)) = 0;
90 ls = ls';
91 ls = ls * (1/sqrt(2));
92
93 erg = mix(lc,ls,3,Fs,fg);
94 erg = erg(1025:length(erg)-(1024+auffuell));
95 wavwrite(erg,Fs,bits,'c:\output\links');
96
97 clear ls lc;
98
99
100 %% Mischung RC-RS
101
102
103 [rc,Fs,bits] = wavread('C:\temp\rc.wav');
104 [rs1,Fs,bits] = wavread(strcat(pfad,'SR.wav'));
105
106 rs = 0;
107 rs(1025:(length(rs1)+1024)) = rs1;
108 clear rs1;
109 rs((length(rs)+1):(length(rs)+auffuell+1024)) = 0;
110 rs = rs';
111 rs = rs * (1/sqrt(2));
112
113 erg = mix(rc,rs,4,Fs,fg);
114 erg = erg(1025:length(erg)-(1024+auffuell));
115 wavwrite(erg,Fs,bits,'c:\output\rechts');
116
117 clear rs rc;
118
119
120 %% Stereo-Output
121
122 % es wird überprüft, ob das Verzeichnis 'encodiert' vorhanden ist
123 speicher_pfad = strcat(verzeichnis,'encodiert\');
124 if (isdir(speicher_pfad) == 0)
125     mkdir(speicher_pfad);
126 end
127
128 % damit's schöner aussieht: Zahlen mit nur einer Ziffer wird eine '0' vorangestellt.
129 Uhr = clock;
130
131 monat = int2str(Uhr(2));
132 if (length(monat)==1)
133     monat = strcat('0',monat);
134 end
135
136 tag = int2str(Uhr(3));
137 if (length(tag)==1)
138     tag = strcat('0',tag);
139 end
140
141 speicher_datei = strcat(speicher_pfad,bezeichnung,int2str(Uhr(1)),monat,tag,'_IRT#5j');
142 wavwrite([wavread('c:\output\links') erg],Fs,bits,speicher_datei);
143
144 clear erg monat tag Uhr speicher_pfad speicher_datei
145
146
147
148
149
150
```

```
151 %% Funktion zum Mischen zweier Signale
152
153 function erg = mix(signal_1,signal_2,mixNr,Fs,fg)
154 % zwei Signale werden mittels des Algorithmus addiert
155
156 % Erzeugung eines Hanningfensters der Länge 2048 Samples
157 fe = hann(2048);
158
159 % Variablen für die Position von Beginn und Ende des Fensters im Audiosignal
160 f1 = 1;
161 f2 = 2048;
162
163 % Startvariable Fensterung
164 f = 0;
165
166 % Festlegung des Ergebnisarrays für die Mischung nach ihrer Bearbeitung
167 erg(1:3072) = 0;
168 erg      = erg';
169
170 % Instanzieren der Fenstervariablen f_signal_1 und f_signal_2
171 f_signal_1 = [0 0];
172 f_signal_2 = [0 0];
173 f_signal_1 = f_signal_1';
174 f_signal_2 = f_signal_2';
175
176
177 % Start der Schleife für die Fensterung
178 while f < fg-1
179
180 % Zero-Padding, Fensterung mit gleichzeitiger Hüllkurvenbewertung durch das
181 % Hanningfenster
182 f_signal_1(1:1024) = 0;
183 f_signal_2(1:1024) = 0;
184 f_signal_1(1025:3072) = signal_1(f1:f2).*fe;
185 f_signal_2(1025:3072) = signal_2(f1:f2).*fe;
186 f_signal_1(3073:4096) = 0;
187 f_signal_2(3073:4096) = 0;
188
189
190 % Durchführung der FFT
191 spektrum_1 = fft(f_signal_1);
192 spektrum_2 = fft(f_signal_2);
193
194
195 % Anpassen der Spektren
196 spektrum_1 = spektrum_ap(spektrum_1);
197 spektrum_2 = spektrum_ap(spektrum_2);
198
199
200 % Bildung des Summensignals nach der Anpassung
201 spektrum_S = spektrum_1+spektrum_2;
202
203
204 % FFT-Koeffizienten werden bearbeitet
205 spektrum_S = koef_bearbeiten(spektrum_1, spektrum_2, spektrum_S, mixNr, f);
206
207
208 % Rücktransformation des bearbeiteten Summensignals in den Zeitbereich
209 signal_S = real(ifft(spektrum_S));
210
211
212 % Zusammenfügen der Fenster:
213 % Aus den gefensterten Blöcken wird hier wieder ein Audiosignal zusammengesetzt.
214 position = length(erg)-3071;
215 erg(position:(length(erg))) = erg(position:length(erg))+signal_S(1:3072);
216 erg(length(erg)+1:length(erg)+1024) = signal_S(3073:4096);
217
218
219 % Position für nächstes Fenster wird festgelegt
220 f1 = f1+1024;
221 f2 = f2+1024;
222
223 f = f+1;
224
225 end
226
227 erg = erg(1025:length(erg)-1024);
228
```

```

229 %% Funktion zur Anpassung der Spektren
230
231 function spektrum = spektrum_ap(spektrum)
232 % Anpassen des Spektrums für spektrum; Nullsetzen des redundanten Anteils
233 laenge = length(spektrum);
234
235 % Die vordere Hälfte des Spektrums wird bis auf den ersten Wert (Gleichanteil) in der
236 % Amplitude verdoppelt
237 spektrum(2:laenge/2) = spektrum(2:laenge/2)*2;
238
239 %Die zweite Hälfte wird gleich null gesetzt.
240 spektrum(laenge/2+1:laenge) = 0;
241
242
243
244 %% Funktion zur Bearbeitung der FFT-Koeffizienten
245
246 function spektrumS = koef_bearbeiten(spektrum1, spektrum2, spektrumS, mixNr, f)
247 % Hier wird eine Fallunterscheidung vorgenommen und je nach Situation eines der drei
248 % Signale an dieser Stelle des Spektrums angepasst
249
250 n=1;
251 while (n <= (length(spektrumS)/2))
252
253 % Einlesen der Amplituden an jeder Stelle der FFT der drei Signale
254     amplitudel = abs(spektrum1(n));
255     amplitud2 = abs(spektrum2(n));
256     ist       = abs(spektrumS(n));
257     soll      = sqrt(amplitudel^2+amplitud2^2);
258
259 % ...sowie der Winkel
260     winkel_ist = angle(spektrumS(n));
261     winkel_1  = angle(spektrum1(n));
262
263 % Fall 1
264     if ist >= 0.5*soll
265         amplitudS_neu = soll+(ist-soll)*0.4;
266         spektrumS(n)  = amplitudS_neu*exp(winkel_ist*i);
267
268 % Fall 2
269     elseif ist < 0.5*soll
270         p = real(spektrum1(n))*real(spektrum2(n))+imag(spektrum1(n))*imag(spektrum2(n));
271         faktor = (-2*p+sqrt(p^2+4*amplitudel^4))/(2*amplitudel^2);
272         faktor = (faktor-1)/1.3+1;
273         amplitudel_neu = amplitudel*faktor;
274         spektrum1(n)   = amplitudel_neu*exp(winkel_1*i);
275         spektrum_neu   = spektrum1(n)+spektrum2(n);
276
277         amp           = abs(spektrum_neu);
278         v             = ist/soll;
279         amplitudS_neu = (v.^2-0.3*v+0.7)*soll+(-(v^2)+0.5*v)*amp;
280         spektrumS(n)  = amplitudS_neu*exp(winkel_ist*i);
281     end
282     n = n+1;
283 end

```

## 13.1.4 Quelltext: IRT-Downmix, Version „IRT\_5r“

```
IRT_5r.m
```

```

1  %% Downmixverfahren IRT (Version IRT_5r)
2  function IRT_5r(verzeichnis,bezeichnung)
3  % IRT-Downmix: Matriziert ein 5.0 Mehrkanalsignal auf ein Stereosignal. Der Additions-
4  % prozess erfolgt mit einer intelligenten Kammfilterkompensation.
5  % function IRT_5r(verzeichnis,bezeichnung)
6  % verzeichnis :Verzeichnisname als String, in dem die Mono-wav-files des
7  % Mehrkanalsignals liegen.
8  % bezeichnung :Bezeichnungskopf der Mono-wav-files des Mehrkanalsignals,
9  %               der bei allen fünf Dateien identisch ist. Ihm folgt die
10 %               Unterscheidung der einzelnen Kanäle mit L,R,C,SL,SR.
11 %IRT_5r.m * Bernfried Runow (Martin Keil) * 03/07/2008
12
13
14
15 %% Mischung L-C
16 % Einlesen der Mehrkanal-Audiosignale im .wav Format
17 pfad = strcat(verzeichnis,bezeichnung);
18 [l1,Fs,bits] = wavread(strcat(pfad,'L.wav'));
19 [c1,Fs,bits] = wavread(strcat(pfad,'C.wav'));
20
21
22 % An den Beginn und ans Ende der eingelesenen Signaldateien wird ein Nullvektor mit
23 % einer Länge von 1024 Samples gefügt. Ebenfalls an das Ende angefügt wird eine Reihe
24 % von Nullen "auffuell", so dass der Vektor genau durch 1024 teilbar wird.
25 auffuell = (fix(length(l1)/1024)+1)*1024-length(l1);
26
27 l = 0;
28 c = 0;
29
30 l(1025:(length(l1)+1024)) = l1;
31 c(1025:(length(c1)+1024)) = c1;
32
33 clear l1 c1;
34
35 l((length(l)+1):(length(l)+auffuell+1024)) = 0;
36 c((length(c)+1):(length(c)+auffuell+1024)) = 0;
37
38
39 % An dieser Stelle wird der Vektor zur weiteren Verarbeitung vom Zeilen ins
40 % Spaltenformat gedreht
41 l = l';
42 c = c';
43
44
45 % Zunächst wird das Centersignal um -3 dB im Pegel abgesenkt
46 c = c * (1/sqrt(2));
47
48
49 % Festlegung, wie oft ein halbes Fenster in den Audiostream passt
50 fg = fix(length(l)/1024);
51
52
53 % Mischung wird durchgeführt
54 erg = mix(l,c,l,Fs,fg);
55
56
57 % Das gemischte Signal wird als wav-file gespeichert
58 wavwrite(erg,Fs,bits,'c:\temp\lc');
59
60 clear l;
61
62
63
64 %% Mischung C-R
65
66 [r1,Fs,bits] = wavread(strcat(pfad,'R.wav'));
67
68 r=0;
69 r(1025:(length(r1)+1024)) = r1;
70 clear r1;
71 r((length(r)+1):(length(r)+auffuell+1024)) = 0;
72 r = r';

```

```

73
74   erg = mix(r,c,2,Fs,fg);
75   wavwrite(erg,Fs,bits,'c:\temp\rc');
76
77   clear r c;
78
79
80
81   %% Mischung LC-LS
82
83   [lc,Fs,bits] = wavread('C:\temp\lc.wav');
84   [ls1,Fs,bits] = wavread(strcat(pfad,'SL.wav'));
85
86   ls = 0;
87   ls(1025:(length(ls1)+1024)) = ls1;
88   clear ls1;
89   ls((length(ls)+1):(length(ls)+auffuell+1024)) = 0;
90   ls = ls';
91   ls = ls * (1/sqrt(2));
92
93   erg = mix(lc,ls,3,Fs,fg);
94   erg = erg(1025:length(erg)-(1024+auffuell));
95   wavwrite(erg,Fs,bits,'c:\output\links');
96
97   clear ls lc;
98
99
100
101   %% Mischung RC-RS
102
103   [rc,Fs,bits] = wavread('C:\temp\rc.wav');
104   [rs1,Fs,bits] = wavread(strcat(pfad,'SR.wav'));
105
106   rs = 0;
107   rs(1025:(length(rs1)+1024)) = rs1;
108   clear rs1;
109   rs((length(rs)+1):(length(rs)+auffuell+1024)) = 0;
110   rs = rs';
111   rs = rs * (1/sqrt(2));
112
113   erg = mix(rc,rs,4,Fs,fg);
114   erg = erg(1025:length(erg)-(1024+auffuell));
115   wavwrite(erg,Fs,bits,'c:\output\rechts');
116
117   clear rs rc;
118
119
120   %% Stereo-Output
121
122   % es wird überprüft, ob das Verzeichnis 'encodiert' vorhanden ist oder noch
123   % erstellt werden muss.
124   speicher_pfad = strcat(verzeichnis,'encodiert\');
125   if (isdir(speicher_pfad) == 0)
126       mkdir(speicher_pfad);
127   end
128
129   % damit's schöner aussieht: Zahlen mit nur einer Ziffer wird eine '0' vorangestellt.
130   Uhr = clock;
131
132   monat = int2str(Uhr(2));
133   if (length(monat)==1)
134       monat = strcat('0',monat);
135   end
136
137   tag = int2str(Uhr(3));
138   if (length(tag)==1)
139       tag = strcat('0',tag);
140   end
141
142   speicher_datei = strcat(speicher_pfad,bezeichnung,int2str(Uhr(1)),monat,tag,'_IRT#5r');
143   wavwrite([wavread('c:\output\links') erg],Fs,bits,speicher_datei);
144
145   clear erg monat tag Uhr speicher_pfad speicher_datei
146
147
148
149
150

```

```

151 %% Funktion zum Mischen zweier Signale
152
153 function erg = mix(signal_1,signal_2,mixNr,Fs,fg)
154 % zwei Signale werden mittels des Algorithmus addiert
155
156 % Erzeugung eines Hanningfensters der Länge 2048 Samples
157 fe = hann(2048);
158
159 % Variablen für die Position von Beginn und Ende des Fensters im Audiosignal
160 f1 = 1;
161 f2 = 2048;
162
163 % Startvariable Fensterung
164 f = 0;
165
166 % Festlegung des Ergebnisarrays für die Mischung nach ihrer Bearbeitung
167 erg(1:3072) = 0;
168 erg      = erg';
169
170 % Instanzieren der Fenstervariablen f_signal_1 und f_signal_2
171 f_signal_1 = [0 0];
172 f_signal_2 = [0 0];
173 f_signal_1 = f_signal_1';
174 f_signal_2 = f_signal_2';
175
176
177 % Start der Schleife für die Fensterung
178 while f < fg-1
179
180 % Zero-Padding, Fensterung mit gleichzeitiger Hüllkurvenbewertung durch das
181 % Hanningfenster
182 f_signal_1(1:1024) = 0;
183 f_signal_2(1:1024) = 0;
184 f_signal_1(1025:3072) = signal_1(f1:f2).*fe;
185 f_signal_2(1025:3072) = signal_2(f1:f2).*fe;
186 f_signal_1(3073:4096) = 0;
187 f_signal_2(3073:4096) = 0;
188
189
190 % Durchführung der FFT
191 spektrum_1 = fft(f_signal_1);
192 spektrum_2 = fft(f_signal_2);
193
194
195 % Anpassen der Spektren
196 spektrum_1 = spektrum_ap(spektrum_1);
197 spektrum_2 = spektrum_ap(spektrum_2);
198
199
200 % Bildung des Summensignals nach der Anpassung
201 spektrum_S = spektrum_1+spektrum_2;
202
203
204 % FFT-Koeffizienten werden bearbeitet
205 spektrum_S = koef_bearbeiten(spektrum_1, spektrum_2, spektrum_S, mixNr, f);
206
207
208 % Rücktransformation des bearbeiteten Summensignals in den Zeitbereich
209 signal_S = real(ifft(spektrum_S));
210
211
212 % Zusammenfügen der Fenster:
213 % Aus den gefensterten Blöcken wird hier wieder ein Audiosignal zusammengesetzt.
214 position = length(erg)-3071;
215 erg(position:(length(erg))) = erg(position:length(erg))+signal_S(1:3072);
216 erg(length(erg)+1:length(erg)+1024) = signal_S(3073:4096);
217
218
219 % Position für nächstes Fenster wird festgelegt
220 f1 = f1+1024;
221 f2 = f2+1024;
222
223 f = f+1;
224
225 end
226
227 erg = erg(1025:length(erg)-1024);
228

```

```
229 %% Funktion zur Anpassung der Spektren
230
231 function spektrum = spektrum_ap(spektrum)
232 % Anpassen des Spektrums für spektrum; Nullsetzen des redundanten Anteils
233 laenge = length(spektrum);
234
235 % Die vordere Hälfte des Spektrums wird bis auf den ersten Wert (Gleichanteil) in der
236 % Amplitude verdoppelt
237 spektrum(2:laenge/2) = spektrum(2:laenge/2)*2;
238
239 %Die zweite Hälfte wird gleich null gesetzt.
240 spektrum(laenge/2+1:laenge) = 0;
241
242
243
244 %% Funktion zur Bearbeitung der FFT-Koeffizienten
245
246 function spektrumS = koef_bearbeiten(spektrum1, spektrum2, spektrumS, mixNr, f)
247 % Hier wird eine Fallunterscheidung vorgenommen und je nach Situation eines der drei
248 % Signale an dieser Stelle des Spektrums angepasst
249
250 n=1;
251 while (n <= (length(spektrumS)/2))
252
253 % Einlesen der Amplituden an jeder Stelle der FFT der drei Signale
254     amplitudel = abs(spektrum1(n));
255     amplituden2 = abs(spektrum2(n));
256     ist = abs(spektrumS(n));
257     soll = sqrt(amplitudel^2+amplituden2^2);
258
259 % ...sowie der Winkel
260     winkel_ist = angle(spektrumS(n));
261     winkel_1 = angle(spektrum1(n));
262
263 % Fall 1
264     if (ist >= soll)
265         amplitudenS_neu = soll+(ist-soll)*0.3;
266         spektrumS(n) = amplitudenS_neu*exp(winkel_ist*i);
267
268 % Fall 2
269     elseif (ist < soll)
270         p = real(spektrum1(n))*real(spektrum2(n))+imag(spektrum1(n))*imag(spektrum2(n));
271         faktor = (-p+sqrt(p^2/4+amplitudel^4))/amplitudel^2;
272         amplituden1_neu = amplitudel*faktor;
273         spektrum1(n) = amplituden1_neu*exp(winkel_1*i);
274         spektrum_neu = spektrum1(n)+spektrum2(n);
275         if (abs(spektrum_neu)<=ist)
276             % Ist-Zustand bleibt erhalten
277         else
278             spektrumS(n)=spektrum_neu;
279         end
280     end
281     n = n+1;
282 end
```

### 13.1.5 Direkter Vergleich der Algorithmen

aus IRT\_5keil.m

```

% Fall 1
if ist < 0.1*soll
    amplitudel_neu = amplitudel*1.9;
    spektrum1(n)   = amplitudel_neu*exp(winkel_1*i);
    spektrumS(n)   = spektrum1(n)+spektrum2(n);

% Fall 2
elseif ist < 0.6*soll && ist >= 0.1*soll
    faktor         = (soll/ist)*0.05+1.4;
    amplitudel_neu = amplitudel*faktor;
    spektrum1(n)   = amplitudel_neu*exp(winkel_1*i);
    spektrumS(n)   = spektrum1(n)+spektrum2(n);

% Fall 4
elseif ist > soll
    amplitudes_neu = soll+(ist-soll)*0.1;
    spektrumS(n)   = amplitudes_neu*exp(winkel_ist*i);

% Tritt keiner der oberen Fälle ein, so erfolgt keine Anpassung (Fall 3)
else
end

```

aus IRT\_5ib.m

```

% Fall 1
if (ist >= soll)
    amplitudes_neu = soll+(ist-soll)*0.1;
    spektrumS(n)   = amplitudes_neu*exp(winkel_ist*i);

% Fall 2
elseif (ist < soll)
    p = real(spektrum1(n))*real(spektrum2(n))+imag(spektrum1(n))*imag(spektrum2(n));
    faktor = (-p+sqrt(p^2/4+amplitudel^4))/amplitudel^2;
    amplitudel_neu = amplitudel*faktor;
    spektrum1(n)   = amplitudel_neu*exp(winkel_1*i);
    spektrum_neu   = spektrum1(n)+spektrum2(n);
    if (abs(spektrum_neu)<=ist)
        % Ist-Zustand bleibt erhalten
    else
        spektrumS(n)=spektrum_neu;
    end
end
end

```

aus IRT\_5j.m

```

% Fall 1
if ist >= 0.5*soll
    amplitudes_neu = soll+(ist-soll)*0.4;
    spektrumS(n)   = amplitudes_neu*exp(winkel_ist*i);

% Fall 2
elseif ist < 0.5*soll
    p = real(spektrum1(n))*real(spektrum2(n))+imag(spektrum1(n))*imag(spektrum2(n));
    faktor = (-2*p+sqrt(p^2+4*amplitudel^4))/(2*amplitudel^2);
    faktor = (faktor-1)/1.3+1;
    amplitudel_neu = amplitudel*faktor;
    spektrum1(n)   = amplitudel_neu*exp(winkel_1*i);
    spektrum_neu   = spektrum1(n)+spektrum2(n);

    amp           = abs(spektrum_neu);
    v              = ist/soll;
    amplitudes_neu = (v.^2-0.3*v+0.7)*soll+(-(v^2)+0.5*v)*amp;
    spektrumS(n)   = amplitudes_neu*exp(winkel_ist*i);
end

```

```
% Fall 1
if (ist >= soll)
    amplitudeS_neu = soll+(ist-soll)*0.3;
    spektrumS(n)   = amplitudeS_neu*exp(winkel_ist*i);

% Fall 2
elseif (ist < soll)
    p = real(spektrum1(n))*real(spektrum2(n))+imag(spektrum1(n))*imag(spektrum2(n));
    faktor = (-p+sqrt(p^2/4+amplitude1^4))/amplitude1^2;
    amplitude1_neu = amplitude1*faktor;
    spektrum1(n)   = amplitude1_neu*exp(winkel_1*i);
    spektrum_neu   = spektrum1(n)+spektrum2(n);
    if (abs(spektrum_neu)<=ist)
        % Ist-Zustand bleibt erhalten
    else
        spektrumS(n)=spektrum_neu;
    end
end
```

### 13.1.6 Quelltext: Funktion zur Darstellung der Bearbeitung der DFT-Koeffizienten

```

                                                                    dft_koef_darstellung.m
1  %% Bearbeitungskennlinie der DFT-Koeffizienten beim IRT-Downmix
2  function dft_koef_darstellung(amplitude)
3  % Die Funktion erstellt zwei idealistische Kennlinien, die Auskunft über die
4  % Bearbeitung der Amplituden beim IRT-Downmix geben. Die Bearbeitungsdaten eines
5  % stochastischen Signals werden hinzukommend in die Grafik integriert.
6  % amplitude :Wert der Amplitude z.B. 1
7  % dft_koef_darstellung.m * Bernfried Runow * 03/07/2008
8
9
10 soll = sqrt(2*amplitude^2);
11
12 %% idealistisch mit konstanter Phasenlage
13 y2 = 0:0.001:amplitude;
14 n=1;
15 while (n<=length(y2))
16     y1(length(y2)+1-n) =y2(n);
17     n=n+1;
18 end
19
20 y2 = y2*exp(0*i);
21 y1 = y1*exp(pi*i);
22 yy = [y1(1:length(y1)-1) y2];
23
24 n=1;
25 while (n<=length(yy))
26     xx(n) = sqrt(2*amplitude^2-yy(n)^2);
27     n=n+1;
28 end
29
30 [spektrumS2 soll_2 ist_2] = berechnen(xx,yy,0.25);
31
32 % Normierung
33 spektrumS2 = abs(spektrumS2)/soll;
34 soll_2      = ones(1,length(soll_2));
35 ist_2      = ist_2/soll;
36
37
38 %% idealistisch mit konstanter Amplitude
39 winkel = 0:0.1:180;
40 y = amplitude*exp(i*winkel*pi/180);
41 x(1:1801) = amplitude;
42
43 [spektrumS soll_ ist_] = berechnen(x,y,0.25);
44
45 % Normierung
46 spektrumS = abs(spektrumS)/soll;
47 soll_     = ones(1,length(soll_));
48 ist_     = ist_/soll;
49
50
51 %% reale Verteilung aus Datei laden
52 datei = load('winkel.mat');
53 [l b] = size(datei.inputV);
54 n = 1;
55 input_v = 0;
56 output_v = 0;
57 while (n<=b)
58     input_v = [input_v datei.inputV(1:l,n)'];
59     output_v = [output_v datei.outputV(1:l,n)'];
60     n = n+8;
61 end
62
63
64 %% Grafik ausgeben
65 FIG1 = figure('PaperType','<custom>','Name','Input/Output-Diagramm',...
66             'Position',[1 1 1482 840],'Renderer','painters');
67
68 axes1 = axes('Parent',FIG1,...
69             'XColor',[0.4 0.4 0.4],'YColor',[0.4 0.4 0.4],...
70             'PlotBoxAspectRatio',[1 1 1],...
71             'XTick',[0.2 0.4 0.6 0.8 1],'XGrid','on',...
72             'FontSize',18);

```

```

73 box('on');
74 hold('all');
75 axis([0 amplitude*2/soll 0 amplitude*2/soll]);
76
77 p_fuell = plot(1,0.5,'Color',[1 1 1]) % Abstandhalter für mehrzeilige Legendeneinträge;
78 p_real = plot(input_v,output_v,'Color',[1 0.85 0],...
79             'MarkerSize',0.5,'Marker','.','LineStyle','none');
80 p_ist = plot(ist_,ist_,'g','LineWidth',4);
81 p_soll = plot(ist_,soll_,'b','LineWidth',4);
82 p_amp = plot(ist_2,pektrumS2,'Color',[1 0.5 0],'LineWidth',4);
83 p_phase = plot(ist_,pektrumS,'Color',[1 0.1 0],'LineWidth',4);
84
85 xlabel('Input DFT-Koeffizient ( {\itIST/SOLL} )','Color',[0 0 0],'FontSize',18);
86 ylabel('Output DFT-Koeffizient','Color',[0 0 0],'FontSize',18);
87
88 legend([p_ist,p_fuell,p_soll,p_fuell,p_amp,p_fuell,p_fuell,p_fuell,p_phase,p_fuell,...
89         p_fuell,p_fuell,p_real,p_fuell],...
90        '{\itIST}',' ',...
91        '{\itSOLL}',' ',...
92        'IRT-Downmix',...
93        '(idealistisch mit konstanter Phase',...
94        'und variabler Amplitude)',' ',...
95        'IRT-Downmix',...
96        '(idealistisch mit konstanter Amplitude',...
97        'und variabler Phase)',' ',...
98        'Reale Verteilung eines stochasti-',...
99        'schen Signals',...
100       'Location','SouthEastOutside');
101
102
103 %% Funktion zur Berechnung der Kurven
104 function [pektrumS soll_ ist_] = berechnen(x,y,kompensation)
105
106 spektrum1 = x;
107 spektrum2 = y;
108 spektrumS = spektrum1+pektrum2;
109 n=1;
110 faktor_ = 0;
111 af = 1;
112
113 while (n<=length(x))
114
115 % Einlesen der Amplituden an jeder Stelle der FFT der drei Signale
116     amplitudel = abs(pektrum1(n));
117     amplitudel2 = abs(pektrum2(n));
118     ist = abs(pektrumS(n));
119     ist_(n) = abs(pektrumS(n));
120     soll = sqrt(amplitudel^2+amplitudel2^2);
121     soll_(n) = soll;
122
123     winkel_ist = angle(pektrumS(1,n));
124     winkel_1 = angle(pektrum1(n));
125
126 % An dieser Stelle kann der Algorithmus eingefügt werden, der untersucht werden soll,
127 % wie zum Beispiel der aus der Version IRT_5ib:
128
129 %% IRT_5ib
130 % Fall 1
131     if (ist >= soll)
132         amplitudelS_neu = soll+(ist-soll)*0.1;
133         pektrumS(n) = amplitudelS_neu*exp(winkel_ist*i);
134
135 % Fall 2
136     elseif (ist < soll)
137         p = real(pektrum1(n))*real(pektrum2(n))+imag(pektrum1(n))*imag(pektrum2(n));
138         faktor = (-p+sqrt(p^2/4+amplitudel^4))/amplitudel^2;
139         amplitudel_neu = amplitudel*faktor;
140         pektrum1(n) = amplitudel_neu*exp(winkel_1*i);
141         pektrum_neu = pektrum1(n)+pektrum2(n);
142         if (abs(pektrum_neu)<=ist)
143             % Ist-Zustand bleibt erhalten
144         else
145             pektrumS(n)=pektrum_neu;
146         end
147     end
148
149     n = n+1;
150 end

```

### 13.1.7 Quelltext: Ausgabe einer Fallanalyse zweier DFT-Koeffizienten

```

                                                                    plot_dft_koef.m
1  %% Ausgabe einer Fallanalyse zweier zu summierender DFT-Koeffizienten
2  function plot_dft_koef
3  % Die Funktion gibt die über die Datei 'dft_wert.mat' eingelesenen
4  % DFT-Koeffizienten grafisch in der komplexen Ebene aus.
5  % plot_dft_koef.m * Bernfried Runow * 03/07/2008
6
7  dft_wert = load('dft_wert.mat');
8  spek1 = dft_wert.spek1;
9  spek2 = dft_wert.spek2;
10 spekS = dft_wert.spekS;
11 spek1_neu = dft_wert.spek1_neu;
12 spek2_neu = dft_wert.spek2_neu;
13 spekS_neu = dft_wert.spekS_neu;
14
15 kein_spek12 = 0;
16
17 % zuerst wird überprüft ob, und wenn ja welcher der beiden zu summierenden
18 % DFT-Koeffizienten bearbeitet wurde.
19 if (spek1_neu~=spek1)
20     spek12_neu = spek1_neu;
21     s_spek12_neu = 'DFT-Koeffizient A (bearbeitet)';
22     farbe = [0.1 0.9 1];
23 elseif (spek2_neu~=spek2)
24     spek12_neu = spek2_neu;
25     s_spek12_neu = 'DFT-Koeffizient B (bearbeitet)';
26     farbe = [0.5 1 0.7];
27 else
28     spek12_neu = 0;
29     kein_spek12 = 1;
30 end
31
32
33 % Die auszugebenden Koeffizienten werden anhand der Größe der Amplitude
34 % sortiert, so dass später ein kleiner Pfeil von einem großen überdeckt
35 % wird
36 a = [abs(spek1) abs(spek2) abs(spekS) abs(spek12_neu) abs(spekS_neu)];
37 a = sort(a);
38
39
40 %% Ausgabe der Grafik
41 FIG = figure('Name','Addition zweier DFT-Koeffizienten',...
42             'PaperType','<custom>', ...
43             'Position',[1 1 1540 800],...
44             'Renderer','painters');
45
46 axes1 = axes('Parent',FIG,...
47             'XColor',[0.4 0.4 0.4],...
48             'YColor',[0.4 0.4 0.4],...
49             'FontSize',18);
50
51 box('on');
52
53 % Zuerst wird der längste Pfeil, zuletzt der kürzeste geplottet.
54 n=5;
55 while (n>0)
56
57     if (abs(spek12_neu)==a(n))
58         if (kein_spek12==0)
59             p_spek12_neu = compass(spek12_neu);
60             set(p_spek12_neu,'LineWidth',4,'Color',farbe);
61         end
62     end
63
64     if (abs(spekS_neu)==a(n))
65         p_spekS_neu = compass(spekS_neu,'m');
66         set(p_spekS_neu,'LineWidth',4);
67     end
68
69     if (abs(spekS)==a(n))
70         p_spekS = compass(spekS,'r');

```

```
71     set(p_spekS, 'LineWidth', 4);
72     end
73
74     if (abs(spek1)==a(n))
75         p_spek1 = compass(spek1, 'b');
76         set(p_spek1, 'LineWidth', 4);
77     end
78
79     if (abs(spek2)==a(n))
80         p_spek2 = compass(spek2, 'g');
81         set(p_spek2, 'LineWidth', 4, 'Color', [0 0.8 0]);
82     end
83
84     hold on;
85     n=n-1;
86
87 end
88
89 % Ausgabe der Legende. Je nach Fall mit bearbeitetem zu summierenden
90 % Koeffizienten oder ohne.
91 if (kein_spek12==0)
92     legend([p_spek1,p_spek2,p_spekS,p_spek12_neu,p_spekS_neu],...
93           'DFT-Koeffizient A (unbearbeitet)',...
94           'DFT-Koeffizient B (unbearbeitet)',...
95           'Summe der beiden DFT-Koeffizienten (unbearbeitet)',...
96           s_spek12_neu, ...
97           'Summe der beiden DFT-Koeffizienten (bearbeitet)',...
98           'Location', 'SouthEastOutside');
99 else
100    legend([p_spek1,p_spek2,p_spekS,p_spekS_neu],...
101          'DFT-Koeffizient A (unbearbeitet)',...
102          'DFT-Koeffizient B (unbearbeitet)',...
103          'Summe der beiden DFT-Koeffizienten (unbearbeitet)',...
104          'Summe der beiden DFT-Koeffizienten (bearbeitet)',...
105          'Location', 'SouthEastOutside');
106
107 end
```

## 13.2 Berechnung der Pegel einer Phantomschallquelle abhängig vom Winkel

Der Summenpegel bei Wiedergabe der 5.0-Mehrkanalversion:

$$P_{5.0} = 10 \cdot \log_{10} \left( 10^{\frac{P_{5.0L}}{10}} + 10^{\frac{P_{5.0C}}{10}} \right) \quad (13.1)$$

Der Summenpegel bei Wiedergabe der Stereoverision:

$$P_{2.0} = 10 \cdot \log_{10} \left( 10^{\frac{P_{2.0L}}{10}} + 10^{\frac{P_{2.0R}}{10}} \right) \quad (13.2)$$

Der Abbildungswinkel der Phantomschallquelle bei Wiedergabe der 5.0-Mehrkanalversion:

$$\vartheta_{5.0} = 15^\circ + \Delta P_{5.0} \cdot 1,1^\circ, \text{ wobei } \Delta P_{5.0} = P_{5.0L} - P_{5.0C}. \quad (13.3)$$

Der Abbildungswinkel der Phantomschallquelle bei Wiedergabe der 5.0-Mehrkanalversion:

$$\vartheta_{2.0} = \Delta P_{2.0} \cdot 2,2^\circ, \text{ wobei } \Delta P_{2.0} = P_{2.0L} - P_{2.0R}. \quad (13.4)$$

Durch Umformung von (13.3) erhält man:

$$\Delta P_{5.0} = \frac{\vartheta_{5.0} - 15^\circ}{1,1^\circ}. \quad (13.5)$$

Gleichsetzen mit der Pegeldifferenz  $\Delta P_{5.0}$  aus (13.3) und (13.5) ergibt:

$$P_{5.0L} - P_{5.0C} = \frac{\vartheta_{5.0} - 15^\circ}{1,1^\circ}$$

$$P_{5.0L} = \frac{\vartheta_{5.0} - 15^\circ}{1,1^\circ} + P_{5.0C} \quad (13.6)$$

(13.6) kann jetzt in (13.1) eingesetzt werden, so dass man folgende Gleichung erhält:

$$P_{5.0} = 10 \cdot \log_{10} \left( 10^{\frac{\frac{\vartheta_{5.0} - 15^\circ}{1,1^\circ} + P_{5.0C}}{10}} + 10^{\frac{P_{5.0C}}{10}} \right) \quad (13.7)$$

Umformen und Auflösen nach  $P_{5.0C}$  führt zum gesuchten Ergebnis

$$\begin{aligned} 10^{\frac{P_{5.0}}{10}} &= 10^{\frac{\frac{\vartheta_{5.0} - 15^\circ}{1,1^\circ} + P_{5.0C}}{10}} + 10^{\frac{P_{5.0C}}{10}} \\ &= 10^{\frac{\vartheta_{5.0} - 15^\circ}{1,1^\circ \cdot 10}} \cdot 10^{\frac{P_{5.0C}}{10}} + 10^{\frac{P_{5.0C}}{10}} \\ &= 10^{\frac{P_{5.0C}}{10}} \cdot \left( 10^{\frac{\vartheta_{5.0} - 15^\circ}{1,1^\circ \cdot 10}} + 1 \right) \end{aligned}$$

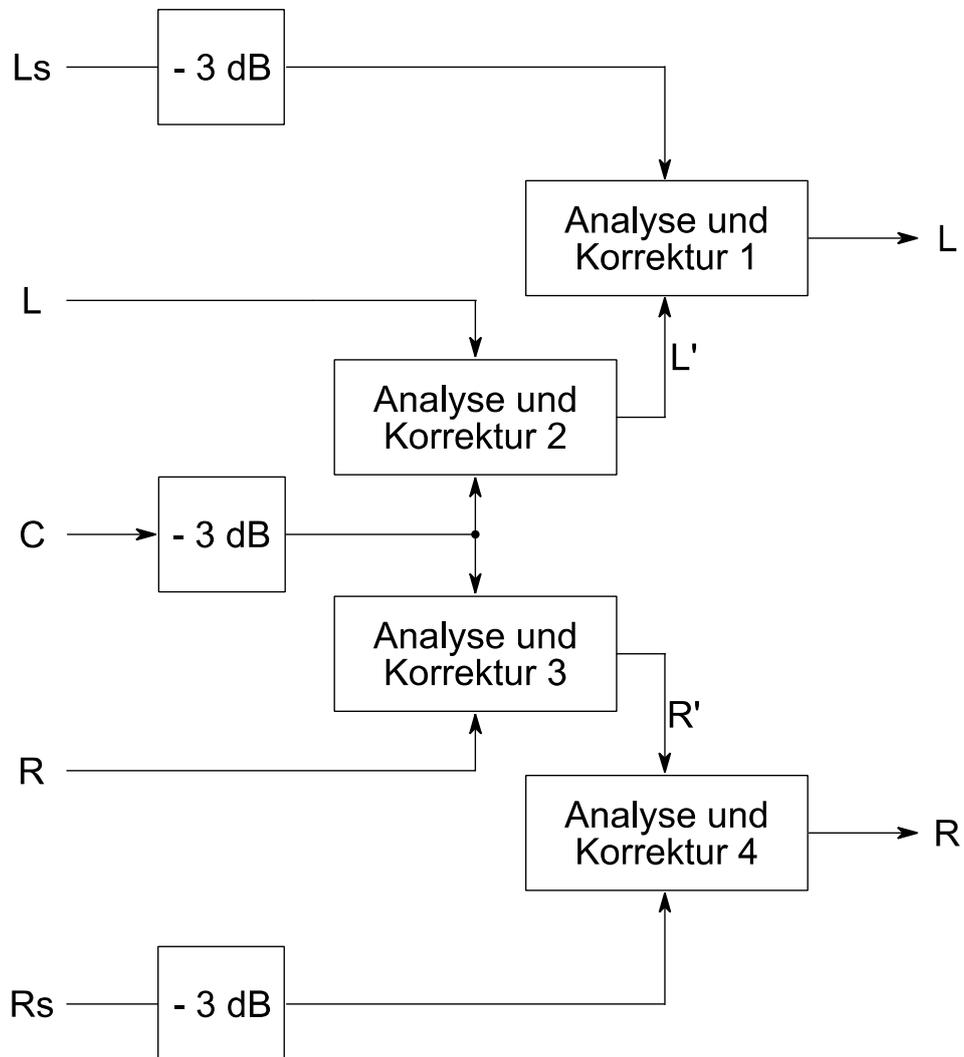
$$10^{\frac{P_{5.0C}}{10}} = \frac{10^{\frac{P_{5.0}}{10}}}{10^{\frac{\vartheta_{5.0}-15^\circ}{1,1^\circ \cdot 10}} + 1}$$

$$P_{5.0C} = 10 \cdot \log_{10} \left( \frac{10^{\frac{P_{5.0}}{10}}}{10^{\frac{\vartheta_{5.0}-15^\circ}{1,1^\circ \cdot 10}} + 1} \right) \quad (13.8)$$

Durch Umformung von (13.1) kann schließlich auch noch  $P_{5.0L}$  berechnet werden.

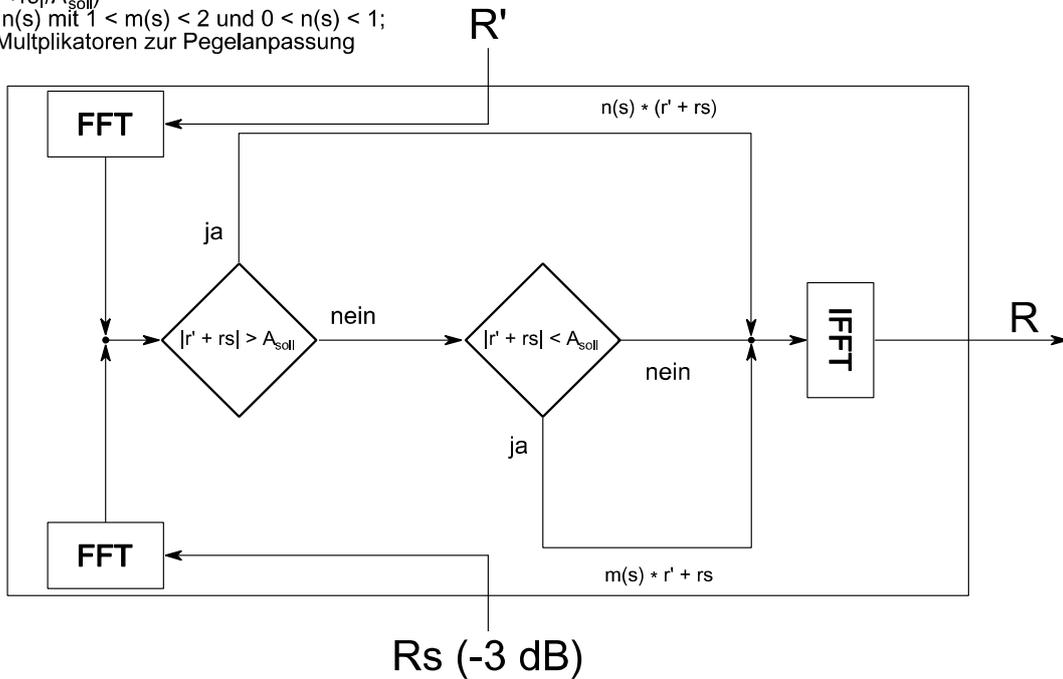
$$P_{5.0L} = 10 \cdot \log_{10} \left( 10^{\frac{P_{5.0}}{10}} - 10^{\frac{P_{5.0C}}{10}} \right) \quad (13.9)$$

### 13.3 Blockschaltbild des IRT-Downmix



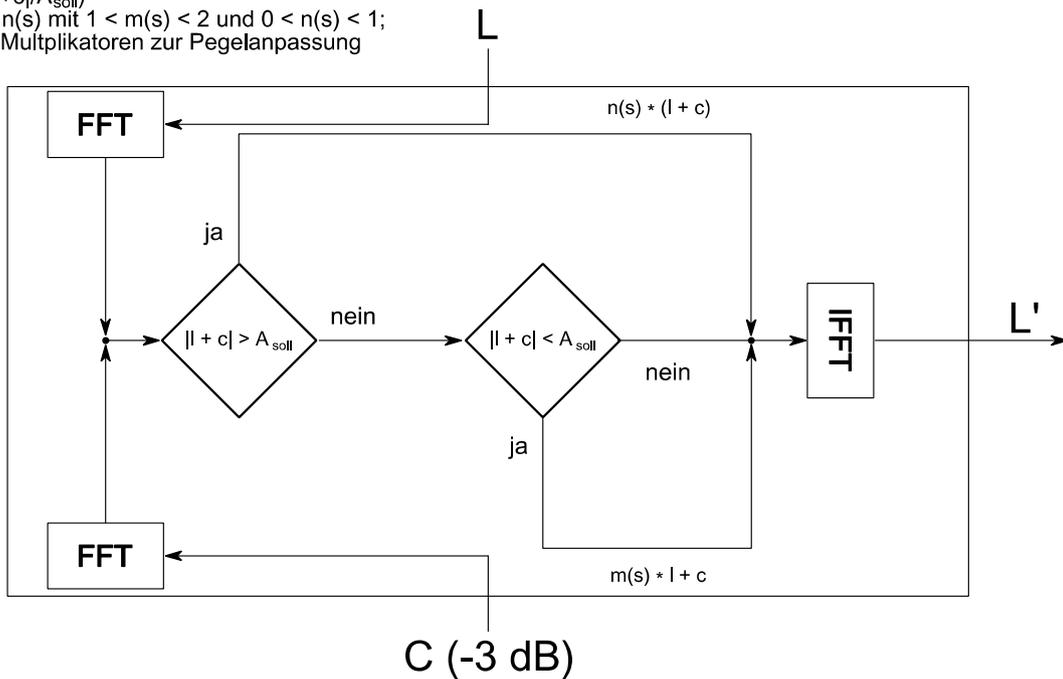
## Analyse und Korrektur 1

$R', R_s \in \mathbb{R}$ , Sample des linken und Center-Kanal in der Zeitdomäne  
 $r', r_s \in \mathbb{C}$ , Sample des linken und Center-Kanal in der Frequenzdomäne  
 $A_{\text{soll}} = \sqrt{(|l|^2 + |c|^2)}$ ; errechneter Sollwert  
 $s = (|r' + r_s| / A_{\text{soll}})$   
 $m(s), n(s)$  mit  $1 < m(s) < 2$  und  $0 < n(s) < 1$ ;  
 Multiplikatoren zur Pegelanpassung



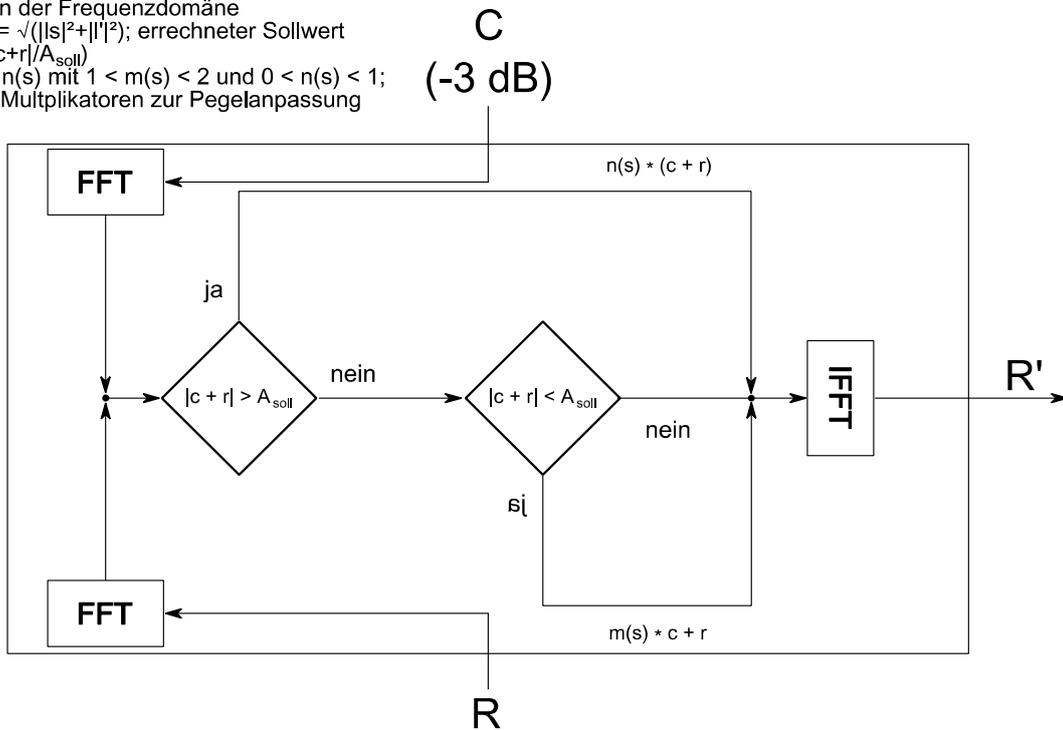
## Analyse und Korrektur 2

$L, C \in \mathbb{R}$ , Sample des linken und Center-Kanal in der Zeitdomäne  
 $l, c \in \mathbb{C}$ , Sample des linken und Center-Kanal in der Frequenzdomäne  
 $A_{\text{soll}} = \sqrt{(|l|^2 + |c|^2)}$ ; errechneter Sollwert  
 $s = (|l + c| / A_{\text{soll}})$   
 $m(s), n(s)$  mit  $1 < m(s) < 2$  und  $0 < n(s) < 1$ ;  
 Multiplikatoren zur Pegelanpassung



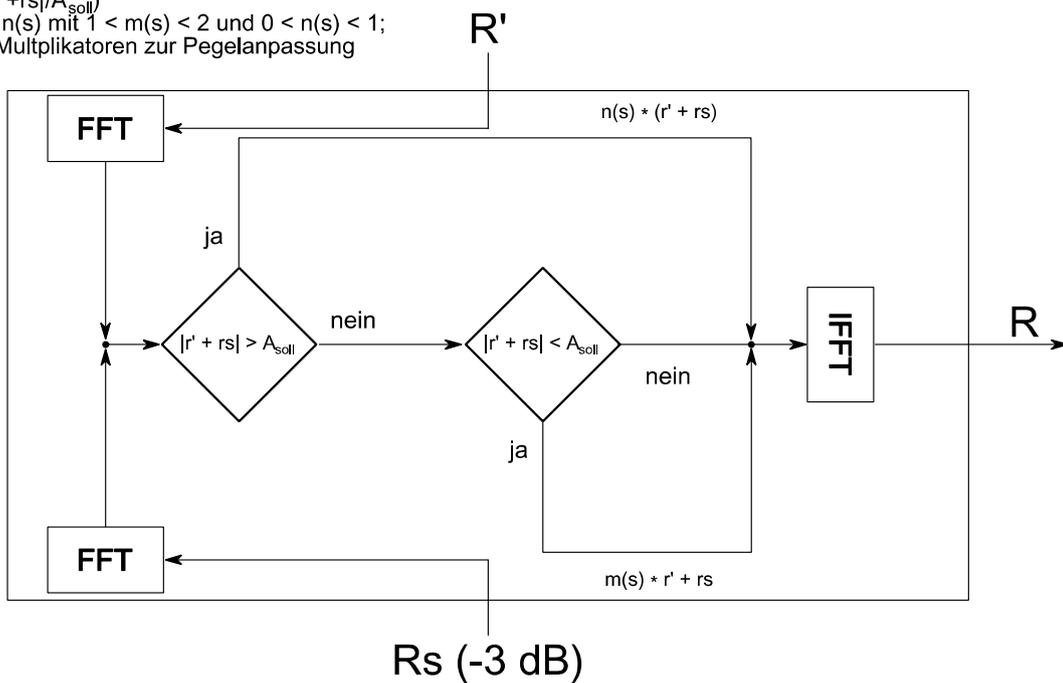
### Analyse und Korrektur 3

$C, R \in \mathbb{R}$ , Sample des linken und Center-Kanal in der Zeitdomäne  
 $c, r \in \mathbb{C}$ , Sample des linken und Center-Kanal in der Frequenzdomäne  
 $A_{\text{soll}} = \sqrt{(|s|^2 + |r|^2)}$ ; errechneter Sollwert  
 $s = (|c+r|/A_{\text{soll}})$   
 $m(s), n(s)$  mit  $1 < m(s) < 2$  und  $0 < n(s) < 1$ ;  
 Multiplikatoren zur Pegelanpassung



### Analyse und Korrektur 4

$R', R_s \in \mathbb{R}$ , Sample des linken und Center-Kanal in der Zeitdomäne  
 $r', r_s \in \mathbb{C}$ , Sample des linken und Center-Kanal in der Frequenzdomäne  
 $A_{\text{soll}} = \sqrt{(|r|^2 + |c|^2)}$ ; errechneter Sollwert  
 $s = (|r'+r_s|/A_{\text{soll}})$   
 $m(s), n(s)$  mit  $1 < m(s) < 2$  und  $0 < n(s) < 1$ ;  
 Multiplikatoren zur Pegelanpassung



## 13.4 Schriftliche Einführung für den Hörversuch

### Einführung

Sie nehmen an einer Evaluierung automatischer Downmixverfahren von 5.1 auf 2.0 teil. In einem Paarvergleich soll die Qualität der zu untersuchenden Zweikanalversionen beurteilt werden.

Das Programm „Step“, welches Sie durch den Hörversuch führt, bietet Ihnen immer zwei Downmixversionen A und B an, die Sie über die Tasten „A“ und „B“ auswählen können. Um zur Orientierung die ursprüngliche Mehrkanalversion zu hören, drücken Sie die Taste „REF“. Mit den Tasten „Play“, „Pause“ und „Loop“ können Sie den Player bedienen - mit den Tasten „Start“ und „Stop“ den Abhörbereich einschränken.

Bewerten Sie immer welche Version - **A oder B** - besser ist. Sehen Sie die 5.1-Version nur als Orientierung für Ihre Bewertung an und **nicht als Bezugspunkt**.

Bewerten Sie letztendlich subjektiv die Qualität der beiden Beispiele auf einer Skala von „A is much better than B“ bis „B is much better than A“.

Folgende Attribute können für die Qualität eines Downmixes entscheidend sein:

Räumlichkeit:	Stimmigkeit der Darstellung des abgebildeten Raumes
Transparenz:	deutliche Differenzierbarkeit der abgebildeten Klangobjekte
Abbildung:	der subjektive Eindruck, dass alle Klangobjekte richtig im Klangbild positioniert und eindeutig wahrnehmbar sind
Klangfarbe:	in diesem Fall die hörbare Änderung zwischen den Testbeispielen durch eine Unter-/Überhöhung bestimmter Frequenzanteile z . B. „Beispiel B klingt dumpfer, als Beispiel A“ oder „Beispiel B klingt spitzer, als Beispiel A“
Lautstärkebalance:	Ausgewogenheit des Lautstärkeverhältnisses in Bezug auf die einzelnen Klangobjekte
Artefaktbildung:	Auftreten von störenden Geräuschen, wie Knackser, Rauigkeiten oder non-lineare Verzerrungen.

Haben Sie eine Session beendet, rufen Sie bitte auf dem nebenstehenden Telefon die Nummer 290 an.

**Bei Problemen und Fragen rufen Sie bitte auch die Nummer 290 an.**

## Quellenverzeichnis

### Literatur

Birkner, Christian (2002): *Surround - Einführung in die Mehrkanalton-Technik*. Bergkirchen: Presse Project Verlags GmbH.

Blackman, R. B., & Tukey, J. W. (1959): Particular Pairs of Window. In *The Measurement of Power Spectra, From the Point of View of Communications Engineering*. New York, Dover.

Blauert, Jens (1974): *Räumliches Hören*. Stuttgart: S. Hirzel Verlag.

Brockhaus (1968): *Der neue Brockhaus* (Bd. 4). Wiesbaden: F. A. Brockhaus.

Deigmöller, Jörg (2006): *Subjektive und objektive Untersuchungen zum 2.0-Downmix von 5.1-Mehrkanal-Produktionen (Diplomarbeit)*. München: Institut für Rundfunktechnik/Fachhochschule Wiesbaden.

Dickreiter, Michael (1997). *Handbuch der Tonstudioteknik* (Bd. 1). (S. f. Rundfunktechnik, Hrsg.) München: K. G. Saur Verlag KG.

Dickreiter, Michael (1997). *Handbuch der Tonstudioteknik* (Bd. 2). (S. f. Rundfunktechnik, Hrsg.) München: K. G. Saur Verlag KG.

Görne, Thomas (2008). *Tontechnik*. (P. D. Schmidt, Hrsg.) München: Carl Hanser Verlag.

Große, Günter (1981). *Von der Edisonwalze zur Stereoplatte - Die Geschichte der Schallplatte*. Berlin: Musikverlag VEB Lied der Zeit.

Grünigen, Daniel Ch. von (2004). *Digitale Signalverarbeitung*. Leipzig: Carl Hanser Verlag.

ITU International Telecommunication Union (1997). *ITU-R BS 1284: General methods for the subjective assessment of sound quality*. Genf: ITU.

ITU International Telecommunication Union. *ITU-R BS.775-2: Multichannel Stereophonic Sound Systems with and without accompanying Picture*. Genf: ITU.

Kammeyer, Karl-Dirk; Kroschel, K. (2006). *Digitale Signalverarbeitung - Filterung und Spektralanalyse mit MATLAB-Übungen*. Wiesbaden: Teubner Verlag.

Keil, Martin (2007). *Realisierung eines Stereo-Downmixes aus einem 5.1 Signal für Anwendungen im Rundfunk (Diplomarbeit)*. München: Institut für Rundfunktechnik/Fachhochschule für Wirtschaft und Technik Berlin.

Naumann, Silvio (2008). *Diplomarbeit: Entwicklung eines Verfahrens zur Generierung, Übertragung und Auswertung lautstärkerelevanter Kennwerte (Diplomarbeit)*. München/Potsdam-Babelsberg: Institut für Rundfunktechnik/Hochschule für Film und Fernsehen "Konrad Wolf".

- Schick, Bastian (2004). *Untersuchung von Downmixverfahren für MP3Surround (Diplomarbeit)*. Detmold, Erlangen: Erich-Thienhaus-Institut der Hochschule für Musik Detmold.
- Schmidt, Ulrich (2003). *Professionelle Videotechnik*. Berlin/Heidelberg: Springer-Verlag.
- Schweizer, Wolfgang (2007). *MATLAB kompakt*. München: Oldenbourg Wissenschaftsverlag GmbH.
- Stoll, Gerhardt (2005). Migration von Stereo zum Mehrkanalton in der Produktion, Distribution und Ausstrahlung. *Institut für Rundfunktechnik - Jahresbericht 2004*, S. 50-53.
- Theile, Günther (2004). *5.1 Surround Sound im Rundfunk - Praxis und State-of-the-Art*. München: IRT-Kolloquium.
- Theile, Günther (1980). *Über die Lokalisation im überlagerten Schallfeld (Dissertation)*. Berlin: Univ.
- Theile, Günther; Krämer, U. (2002). Empfehlungen für die Praxis SSF. *Hörbedingungen und Wiedergabeanordnungen für Mehrkanal-Stereofonie. Surround Sound Forum*.
- Veit, Ivar (2005). *Technische Akustik*. Würzburg: Vogel Industrie Medien GmbH & Co.
- Weltner, Klaus; Wiesner, Hartmut; Heinrich, P.-B.; Engelhardt, P.; Schmidt, H. (2001). *Mathematik für Physiker 1 - Basiswissen für das Grundstudium der Experimentalphysik* (Bd. 1). (K. Weltner, Hrsg.) Berlin: Springer Verlag.
- Wendemuth, Andreas (2005). *Grundlagen der digitalen Signalverarbeitung*. Berlin/Heidelberg: Springer-Verlag.
- Werner, Martin (2006). *Digitale Signalverarbeitung mit MATLAB*. Wiesbaden: Friedr. Vieweg & Sohn Verlag.
- Wittek, Helmut (2000). *Untersuchung zur Richtungsabbildung mit L-C-R Hauptmikrofonen (Diplomarbeit)*. Düsseldorf: FH.
- Zwicker, Eberhard; Feldtkeller, R. (1967). *Das Ohr als Nachrichtenempfänger*. Stuttgart: Hirzel.

## elektronische Quellen

*Digital Audio Broadcasting*. (14. 06. 2008). Von Wikipedia:  
[http://de.wikipedia.org/wiki/Digital\\_Audio\\_Broadcasting](http://de.wikipedia.org/wiki/Digital_Audio_Broadcasting)

*Digital Video Broadcasting*. (24. 06. 2008). Von Wikipedia:  
[http://de.wikipedia.org/wiki/Digital\\_Video\\_Broadcasting](http://de.wikipedia.org/wiki/Digital_Video_Broadcasting)

Griesinger, D. (2003). *Surround From Stereo*. Von David Griesinger's homepage:  
[http://world.std.com/~griesngr/surround\\_from\\_stereo2.ppt](http://world.std.com/~griesngr/surround_from_stereo2.ppt)

*Overdub*. (17. 06. 2008). Von Wikipedia:  
<http://de.wikipedia.org/wiki/Overdub>

*Quadrophonie*. (29. 05. 2008). Von Wikipedia:  
<http://de.wikipedia.org/wiki/Quadrophonie>

Sandner, Prof. Michael (2003). *Surround im Rundfunk*.  
[http://www.m-sandner.de/lehre/VL\\_pdf/Surround\\_im\\_Rundfunk.pdf](http://www.m-sandner.de/lehre/VL_pdf/Surround_im_Rundfunk.pdf)

Sengpiel, E. (07 2003). *Addition von Amplituden und Pegeln*. Von Forum für Mikrofonaufnahmen und Tonstudioteknik:  
<http://www.sengpielaudio.com/AdditionVonAmplitudenUndPegeln.pdf>

*Sequenzen (Musik) - Wikipedia*. (22. 05. 2008). Von Wikipedia:  
[http://de.wikipedia.org/wiki/Sequenzen\\_%28Musik%29](http://de.wikipedia.org/wiki/Sequenzen_%28Musik%29)

*Stereophonic sound*. (16. 06. 2008). Von Wikipedia:  
<http://en.wikipedia.org/wiki/Stereo>

*Super Audio Compact Disc*. (13. 06. 2008). Von Wikipedia:  
[http://de.wikipedia.org/wiki/Super\\_Audio\\_Compact\\_Disc](http://de.wikipedia.org/wiki/Super_Audio_Compact_Disc)

Wittek, Helmut; Riffeser, V. (2005). *Downmix Web-Slides*. Von  
[http://www.hauptmikrofon.de/DownmixDemo/Downmix\\_WebSlides.zip](http://www.hauptmikrofon.de/DownmixDemo/Downmix_WebSlides.zip)

## Abbildungsverzeichnis

Abbildung 2.1-1 Empfehlung der Lautsprecheraufstellung und Abhörposition für Mono nach ITU-R BS.1116 .....	4
Abbildung 2.1-2 Empfehlung der Lautsprecheraufstellung und Abhörposition für Zweikanal-Stereo nach ITU-R BS.1116 ...	5
Abbildung 2.1-3 Empfehlung für die Lautsprecheranordnung und Abhörposition von 3/2-Mehrkanalton nach ITU-R BS.775-2 ....	6
Abbildung 2.1-4 Möglichkeiten der Anordnung von Phantomschallquellen in 2.0-Stereo oder 5.0-Mehrkanal.....	6
Abbildung 3.2-1: Blockschaltbild des Downmix nach ITU BS.775 .....	12
Abbildung 3.2-2 Blockschaltbild des Dolby Pro Logic II Encoders .....	12
Abbildung 3.3-1 Blockschaltbild Logic7 .....	14
Abbildung 4.3-1 Addition zweier kohärenter und gegenseitig unverzögerter Amplituden .....	18
Abbildung 4.3-2 Addition inkohärenter Amplituden, wobei $a \neq b$ .....	19
Abbildung 4.4-1 Entstehung von Notches .....	20
Abbildung 4.4-2 Entstehung von Peaks .....	21
Abbildung 4.5-1 Relative Auslenkung $\Phi$ einer Phantomschallquelle in Abhängigkeit von der Pegeldifferenz $\Delta L$ .....	22
Abbildung 4.5-2 Relative Auslenkung $\Phi$ einer Phantomschallquelle in Abhängigkeit von der Laufzeitdifferenz $\Delta t$ .....	23
Abbildung 4.6-1 Kopfbezogenes Koordinatensystem.....	24
Abbildung 4.7-1 komplexe Zahlenebene .....	24
Abbildung 4.8-1 Eingangs- und Ausgangssignal eines signalverarbeitenden Systems.....	25
Abbildung 4.8-2 Grafische Darstellung einer Faltung $x(t) * h(t) = y(t)$ .....	29
Abbildung 5.1-1 Kammfilter nach einem ITU-Downmix bei $\Delta tR/RS = 0,25 \text{ ms}$ .....	30
Abbildung 5.1-2 Kammfilter nach einem Dolby Pro Logic II-Downmix bei $\Delta tR/RS = 0,25 \text{ ms}$ .....	31
Abbildung 5.1-3 Kammfilter nach einem C.T.-Downmix bei $\Delta tR/RS = 0,25 \text{ ms}$ .....	31
Abbildung 5.2-1 5.0 Abhörsituation mit $L = C = -6 \text{ dBFS}$ .....	32
Abbildung 5.2-2 Ein- und Ausgangspegel nach einem ITU-Downmix mit $C = L = -6 \text{ dBFS}$ .....	33
Abbildung 5.2-3 2.0 Abhörsituation nach einem ITU-Downmix mit $L = C = -6 \text{ dBFS}$ .....	33
Abbildung 5.2-4 5.0 Abhörsituation mit $L = LS = -6 \text{ dBFS}$ .....	33
Abbildung 5.2-5 Ein- und Ausgangspegel nach einem ITU-Downmix mit $L = LS = -6 \text{ dBFS}$ .....	34
Abbildung 5.2-6 2.0 Abhörsituation nach einem ITU-Downmix mit $L = LS = -6 \text{ dBFS}$ .....	34
Abbildung 5.2-7 5.0 Abhörsituation mit $LS = RS = -6 \text{ dBFS}$ .....	34
Abbildung 5.2-8 Ein- und Ausgangspegel nach einem ITU-Downmix mit $LS = RS = -6 \text{ dBFS}$ .....	35
Abbildung 5.2-9 2.0 Abhörsituation nach einem ITU-Downmix mit $LS = RS = -6 \text{ dBFS}$ .....	35
Abbildung 5.2-10 Pegelzunahme nach einem Downmix in Abhängigkeit vom Abbildungswinkel einer Phantomschallquelle in der Mehrkanalversion. ....	36
Abbildung 5.3-1 5.0 Abhörsituation mit einem Differenzpegel von $PDP 5.0 = 2 \text{ dB}$ .....	37
Abbildung 5.3-2 2.0 Abhörsituation mit einem Differenzpegel von $PDP 2.0 = 4,3 \text{ dB}$ .....	37
Abbildung 5.4-1 5.0 Abhörsituation mit Phantomschallquelle bei $L = -6 \text{ dBFS}$ und $C = -6 \text{ dBFS}$ .....	38
Abbildung 5.4-2 2.0 Abhörsituation nach einem ITU-Downmix mit $L = -1,35 \text{ dBFS}$ und $R = -9 \text{ dBFS}$ .....	38
Abbildung 5.4-3 5.0 Abhörsituation mit Phantomschallquelle bei $L = -14,5 \text{ dBFS}$ und $C = -3,5 \text{ dBFS}$ .....	39
Abbildung 5.4-4 2.0 Abhörsituation nach einem ITU-Downmix mit $L = -3,6 \text{ dBFS}$ und $R = -6,5 \text{ dBFS}$ .....	39
Abbildung 5.4-5 Änderung des Abbildungswinkels einer Phantomschallquelle nach einem Downmix in Bezug auf die Mehrkanalversion. Ein positives Vorzeichen entspricht einer Verschiebung nach außen, ein negatives Vorzeichen einer Verschiebung nach innen zur Mitte hin. ....	40
Abbildung 5.4-6 Abbildungsänderung nach einem Dolby Pro Logic II Downmix bei Laufzeitstereofonie.....	41
Abbildung 6.1-1 Grafische Darstellung der Veränderung der DFT-Koeffizienten im Spektralbereich zur Kammfilterkompensation.....	43
Abbildung 6.1-2 Grafische Darstellung der energetischen Summe $SOLL = A[k]^2 + B[k]^2$ .....	43
Abbildung 6.1-3 Grafische Darstellung der Bearbeitung der Summe. ....	44
Abbildung 6.1-4 Grafische Darstellung der Bearbeitung einer der beiden zu summierenden DFT-Koeffizienten, so dass eine Auslöschung in der Summe verhindert wird. ....	44
Abbildung 6.2-1 Blockschaltbild IRT-Downmix.....	46
Abbildung 6.2-2 Bewertung eines gefensterten Audiosignals mit einem Hanning-Fenster. ....	47
Abbildung 6.2-3 Blockdarstellung des IRT-Downmixverfahrens nach Keil. ....	48
Abbildung 6.2-4 Darstellung der Bearbeitung des IRT-Downmixes in der komplexen Ebene bei Fall 1 .....	50

Abbildung 6.2-5 Darstellung der Bearbeitung des IRT-Downmixes in der komplexen Ebene bei Fall 2. ....	50
Abbildung 6.2-6 Darstellung der Bearbeitung des IRT-Downmixes in der komplexen Ebene bei Fall 4. ....	51
Abbildung 6.2-7 Modulare Darstellung des IRT-Downmixes .....	52
Abbildung 7.2-1 Fenster aus dem Input-Signal $c[n] = \sin(2\pi \cdot 200\text{Hz} \cdot nT)$ mit einer Abtastfrequenz von $f_s = 44,1 \text{ kHz}$ nach der Hüllkurvenbewertung.....	55
Abbildung 7.2-2 Fenster mit allen sechs Signalproben. Ergebnis der Eingangssignale $l[n] = \sin(2\pi \cdot 30\text{Hz} \cdot nT)$ und $c[n] = \sin(2\pi \cdot 100\text{Hz} \cdot nT)$ mit einer Abtastfrequenz von $f_s = 44,1 \text{ kHz}$ .....	55
Abbildung 7.2-3 Fenster der beiden Signale $l[n] = 0,4 \cdot \sin 2\pi \cdot 30\text{Hz} \cdot nT$ und $cn = 0,4 \cdot 2 \cdot \sin 2\pi \cdot 30\text{Hz} \cdot nT + \pi$ aus dem IRT-Downmix nach Martin Keil (IRT_5keil). Die Bearbeitung der DFT-Koeffizienten beginnt mit dem zweiten Koeffizienten.....	56
Abbildung 7.2-4 Fenster der beiden Signale $l[n] = 0,4 \cdot \sin 2\pi \cdot 30\text{Hz} \cdot nT$ und $c[n] = 0,4 \cdot 2 \cdot \sin 2\pi \cdot 30\text{Hz} \cdot nT$ aus dem IRT-Downmix nach Martin Keil (IRT_5keil). Die Bearbeitung der DFT-Koeffizienten beginnt mit dem zweiten Koeffizienten.....	57
Abbildung 7.2-5 Fenster der beiden Signale $l[n] = 0,4 \cdot \sin 2\pi \cdot 30\text{Hz} \cdot nT$ und $c[n] = 0,4 \cdot 2 \cdot \sin 2\pi \cdot 30\text{Hz} \cdot nT + \pi$ aus dem IRT-Downmix nach Martin Keil (IRT_5keil). Die Bearbeitung der DFT-Koeffizienten beginnt mit dem ersten Koeffizienten. ....	58
Abbildung 7.2-6 Fenster aus dem IRT-Downmix nach Martin Keil (IRT_5keil). Die Bearbeitung in der Funktion „koef_bearbeiten“ beginnt mit dem ersten Koeffizienten. Das Testmaterial stammt aus der Matthäus-Passion (BWV 244) von J. S. Bach. Darstellung aller sechs Signalproben. ....	59
Abbildung 7.2-7 Fenster aus dem IRT-Downmix nach Martin Keil (IRT_5keil). Die Bearbeitung in der Funktion „koef_bearbeiten“ beginnt mit dem ersten Koeffizienten. Das Testmaterial stammt aus der Matthäus-Passion (BWV 244) von J. S. Bach. Darstellung des Output-Signals. ....	60
Abbildung 7.2-8 Fenster aus dem IRT-Downmix. Die Nullen des „Zero-Paddings“ werden zur Hälfte jeweils vor und nach dem gefensterten Signal angesetzt. Das Testmaterial stammt aus der Matthäus-Passion (BWV 244) von J. S. Bach. Darstellung des ersten Input-Signals. ....	61
Abbildung 7.2-9 Fenster aus dem IRT-Downmix. Die Nullen des „Zero-Paddings“ werden zur Hälfte jeweils vor und nach dem gefensterten Signal angesetzt. Die Bearbeitung in der Funktion „koef_bearbeiten“ beginnt mit dem ersten Koeffizienten. Das Testmaterial stammt aus der Matthäus-Passion (BWV 244) von J. S. Bach. Darstellung der beiden bewerteten Eingangssignale sowie des Ausgangssignals. ....	62
Abbildung 7.3-1 Input-/Outputdiagramm mit den Kennlinien für SOLL, IST und einem fiktiven Downmixalgorithmus mit Kammfilterkompensation.....	63
Abbildung 7.3-2 Testsignalreihen $A1[k]$ (blau) und $B1[k]$ (grün), mit deren Hilfe die Kennlinie des IRT-Downmix berechnet wird. Der weiße Pfeil mit roter Kontur stellt die Summe der beiden Signale dar. Die Entwicklung der Reihen basiert auf festen Amplituden $A1[k] = B1[k]$ und einem variablen Phasenwinkel $0^\circ \leq \angle A1k \leq 180^\circ$ .....	64
Abbildung 7.3-3 Testsignalreihen $A2[k]$ (blau) und $B2[k]$ (grün), mit deren Hilfe die Kennlinie des IRT-Downmix berechnet wird. Der weiße Pfeil mit roter Kontur stellt die Summe der beiden Signale dar. Die Entwicklung der Reihen basiert auf variablen Amplituden und einer konstanten Phasenlage. ....	64
Abbildung 7.3-4 Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix nach Martin Keil „IRT_5keil“ .....	65
Abbildung 7.3-5 Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix nach Martin Keil. Beschriftung der Fälle unterschiedlicher Bearbeitung. ....	66
Abbildung 7.3-6 Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix. Die Bearbeitung erfolgt mit einer Fallunterscheidung. Die Bearbeitung in Fall 1 ( $IST < SOLL$ ) erfolgt mit dem Korrekturfaktor $x[p]$ aus (7.11).....	69
Abbildung 7.3-7 Darstellung der an einem Additionsvorgang des IRT-Downmixes beteiligten DFT-Koeffizienten in der komplexen Ebene. Die Bearbeitung des DFT-Koeffizienten A erfolgt mit dem Korrekturfaktor $p = 6,75$ , der mit Hilfe von (7.11) errechnet wird. Der Phasenwinkel wird radial, die Amplitude zentrifugal dargestellt. $A[r] = 0,029 \cdot e^{0,84 \cdot i}$ und $Br = 0,1 \cdot e^{3,69 \cdot i}$ .....	69
Abbildung 7.3-8 Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix „IRT_5i“. Die Bearbeitung erfolgt mit einer Fallunterscheidung. Die Korrektur in Fall 1 ( $IST < SOLL$ ) erfolgt mit dem Korrekturfaktor $x[p]$ aus (7.12), wobei $w = 0,25$ .....	70

Abbildung 7.3-9 Darstellung der an einem Additionsvorgang des IRT-Downmixes nach Martin Keil beteiligten DFT-Koeffizienten in der komplexen Ebene. Der Phasenwinkel wird radial, die Amplitude zentrifugal dargestellt. $A[r] = 0,0089 \cdot e^{5,17 \cdot i}$ und $Br = 0,0159 \cdot e^{2,05 \cdot i}$ .....	72
Abbildung 7.3-10 Darstellung der an einem Additionsvorgang des IRT-Downmixes beteiligten DFT-Koeffizienten in der komplexen Ebene. Der Algorithmus bearbeitet den DFT-Koeffizienten mit der größeren Amplitude. Der Phasenwinkel wird radial, die Amplitude zentrifugal dargestellt. $A[r] = 0,0089 \cdot e^{5,17 \cdot i}$ und $Br = 0,0159 \cdot e^{2,05 \cdot i}$ .....	73
Abbildung 7.3-11 Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix. Die Bearbeitung erfolgt mit einer Fallunterscheidung bei $IST/SOLL = 1$ . Die Bearbeitung in Fall 1 ( $IST < SOLL$ ) erfolgt mit dem Korrekturfaktor $x[p]$ , wobei $p[k] = ReA[k]ReB[k] + ImA[k]ImB[k]$ . Der DFT-Koeffizient mit der größeren Amplitude wird bearbeitet .....	74
Abbildung 7.3-12 Darstellung der an einem Additionsvorgang des IRT-Downmixes beteiligten DFT-Koeffizienten in der komplexen Ebene. Der Algorithmus bearbeitet den DFT-Koeffizienten mit der größeren Amplitude. Der Phasenwinkel wird radial, die Amplitude zentrifugal dargestellt. $A[r] = 0,48 \cdot e^{1,59 \cdot i}$ und $Br = 0,51 \cdot e^{4,39 \cdot i}$ .....	75
Abbildung 7.3-13 Darstellung der an einem Additionsvorgang des IRT-Downmixes beteiligten DFT-Koeffizienten in der komplexen Ebene. Der Algorithmus bearbeitet den DFT-Koeffizienten mit der größeren Amplitude. Der Phasenwinkel wird radial, die Amplitude zentrifugal dargestellt. $A[r] = 0,51 \cdot e^{1,59 \cdot i}$ und $Br = 0,48 \cdot e^{4,39 \cdot i}$ .....	75
Abbildung 7.3-14 Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix „5ib“. Die Bearbeitung erfolgt mit einer Fallunterscheidung. Die Korrektur in Fall 1 ( $IST < SOLL$ ) erfolgt mit dem Korrekturfaktor $x[p]$ aus (7.12), wobei $w = 0,25$ . Für DFT-Koeffizienten, die mit der Bearbeitung eine Dämpfung erfahren, wird der zugehörige $IST$ -Wert verwendet. ....	76
Abbildung 7.3-15 Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix. Die Bearbeitung erfolgt mit einer Fallunterscheidung bei $IST/SOLL = 0,5$ . Die Bearbeitung in Fall 1 erfolgt mit dem Korrekturfaktor $x[p]$ , wobei $p[k] = ReA[k]ReB[k] + ImA[k]ImB[k]$ .....	77
Abbildung 7.3-16 Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix „5j“. Die Bearbeitung erfolgt mit einer Fallunterscheidung bei $IST/SOLL = 0,5$ , und im 1. Fall in zwei Durchläufen. Die Bearbeitung in Fall 2 erfolgt bereits anhand der Optimierung in Kapitel 8. ....	78
Abbildung 8.1-1 Pegelzunahme nach einem Downmix in Abhängigkeit vom Abbildungswinkel einer Phantomschallquelle in der Mehrkanalversion. ....	81
Abbildung 8.1-2 Änderung des Abbildungswinkels einer Phantomschallquelle nach einem Downmix in Bezug auf die Mehrkanalversion. Ein positives Vorzeichen entspricht einer Verschiebung nach außen, ein negatives Vorzeichen einer Verschiebung nach innen zur Mitte hin. ....	82
Abbildung 9.1-1 Oberfläche des Programms „STEP“ .....	84
Abbildung 9.2-1 Downmix "IRT_5ib" im Vergleich zu "IRT_5keil" .....	88
Abbildung 9.2-2 Downmix "IRT_5ib" im Vergleich zu "ITU" .....	89
Abbildung 9.2-3 Downmix "IRT_5j" im Vergleich zu "ITU" .....	90
Abbildung 9.2-4 Downmix "IRT_5ib" im Vergleich zu "IRT_5j" .....	90
Abbildung 9.2-5 Downmix "IRT_5ib" im Vergleich zum Handmix .....	91
Abbildung 9.3-1 Kennlinien der Bearbeitung der DFT-Koeffizienten mit dem IRT-Downmix „5r“. Die Bearbeitung erfolgt mit einer Fallunterscheidung. Die Korrektur in Fall 1 ( $IST < SOLL$ ) erfolgt mit dem Korrekturfaktor $x[p]$ aus (7.12), wobei $w = 0,25$ . Für DFT-Koeffizienten, die mit der Bearbeitung eine Dämpfung erfahren, wird der zugehörige $IST$ -Wert verwendet. Die Bearbeitung in Fall 2 erfolgt anhand der Optimierung in Kapitel 8. ....	92
Abbildung 10.1-1 Verdeckungsprobleme beim automatischen Downmix von 5.0 auf 2.0 am Beispiel einer Film-mischung .....	94
Abbildung 10.2-1 Prinzip der MPEG-Surround Codierung.....	95
Abbildung 10.2-2 Modell zur Realisierung eines 5.1-Radioprogramms .....	96
<b>Tabellen</b>	
Tabelle 5.4-1 Lokalisationsunschärfe einer durch zwei Lautsprecher erzeugten Phantomschallquelle .....	40

## Glossar

A	AC-3 AES/EBU AMS ARD  Atmo	Audio Coding 3 Algorithm Audio Engineering Society / European Broadcast Union Audio-Massenspeicher Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland Atmosphäre
B	BR	Bayerischer Rundfunk
C	CD Codec  Consumer C.T.	Compact Disc Verfahren zum codieren und decodieren von Signalen oder Daten engl.: Verbraucher Coding Technologies
D	D.A.I.S. DAB dB dBFS dB SPL DFT DMB DolbyE  DPL2 DRM DSV DTRS DTS DVB DVB-T/ -S/ -C/ -H  DVD	Digital Audio Interconnection System Digital Audio Broadcast Dezibel Dezibel Full Scale Dezibel Sound Pressure Level Diskrete Fourier Transformation Digital Multimedia Broadcasting Professionelles Codiersystem für 5.1-Audiosignale in digitaker Zweikanalinfrastuktur Dolby Pro Logic 2 Digital Radio Mondiale Digitale Signalverarbeitung Digital Tape Recording System Digital Theatre System Digital Video Broadcast Digital Video Broadcast-Terrestrial/-Satellite/ -Cable/ -Handhelds Digital Versatile Disc
F	FFT FIR	Fast Fourier Transformation Finite Impuls Response
H	HDTV HM Hochpass  Hz	High Definition Television Handmix Filter, bei dem Frequenzen unter einer bestimmten Grenzfrequenz abgesenkt werden Hertz
I	IC IFA ILD IRT	Interaural Correlation Internationale Funkausstellung Interaural Level Difference Institut für Rundfunktechnik

	ITD	Interaural Time Difference
	ITU	International Telecommunication Union
K		
	kBit/s	Kilo-Bit pro Sekunde
L		
	L/ R / C / Ls/ Rs	Links / Rechts / Center / Links Surround / Rechts Surround
	LO/RO	L-Null / R-Null
	LFE	Low Frequency Effect / Enhancement
	Lt/Rt	Left total / Right total
	LTI	Linear Time-Invariant Systems
M		
	MATLAB	MATrix LABoratory
	MPEG	Moving Picture Expert Group
	MS	Mitte/Seite
O		
	ORF	Österreichischer Rundfunk
P		
	PCM	Pulse Code Modulation
	Plug-in	Erweiterungsmodul für eine Computersoftware
S		
	SACD	Super Audio Compact Disc
	Samplerate	engl.: Abtastfrequenz
	SDTV	Standard Definition Television
	STB	Set Top Box
	SWR	Südwestrundfunk
T		
	Tiefpass	Filter, bei dem Frequenzen über einer bestimmten Grenzfrequenz abgesenkt werden
V		
	VST	Virtual Studio Technologie
W		
	wav	wave
	WDR	Westdeutscher Rundfunk
	Wiesn	Oktoberfest in München
X		
	XY	Mikrofonanordnung
Z		
	ZDF	Zweites Deutsches Fernsehen

## Trackliste der DVD-Audio

Auf der beiliegenden DVD-Audio befinden sich die im Hörversuch verwendeten Hörbeispiele mit den verschiedenen Downmixversionen. Im Ordner „MATLAB Quelltext“ ist außerdem der originale Quelltext abgelegt.

Hörbeispiel	Track	Version
Hörbeispiel 1: Chor	1	Mehrkanal
	2	Downmix IRT_5ib
	3	Downmix IRT_5j
	4	Handmix
	5	Downmix ITU
	6	Downmix IRT_5keil
Hörbeispiel 2: Klassik1	7	Mehrkanal
	8	Downmix IRT_5ib
	9	Downmix IRT_5j
	10	Handmix
	11	Downmix ITU
Hörbeispiel 3: TV-Show	12	Downmix IRT_5keil
	13	Mehrkanal
	14	Downmix IRT_5ib
	15	Downmix IRT_5j
	16	Handmix
Hörbeispiel 4: Hörspiel	17	Downmix ITU
	18	Downmix IRT_5keil
	19	Mehrkanal
	20	Downmix IRT_5ib
	21	Downmix IRT_5j
	22	Handmix
Hörbeispiel 5: Klassik2	23	Downmix ITU
	24	Downmix IRT_5keil
	25	Mehrkanal
	26	Downmix IRT_5ib
	27	Downmix IRT_5j
	28	Handmix
Hörbeispiel 6: Jazz1	29	Downmix ITU
	30	Downmix IRT_5keil
	31	Mehrkanal
	32	Downmix IRT_5ib
	33	Downmix IRT_5j
	34	Handmix
	35	Downmix ITU
	36	Downmix IRT_5keil

Hörbeispiel	Track	Version
Hörbeispiel 7: Jazz2	37	Mehrkanal
	38	Downmix IRT_5ib
	39	Downmix IRT_5j
	40	Handmix
	41	Downmix ITU
	42	Downmix IRT_5keil
Hörbeispiel 8: Orgel	43	Mehrkanal
	44	Downmix IRT_5ib
	45	Downmix IRT_5j
	46	Handmix
	47	Downmix ITU
	48	Downmix IRT_5keil
Hörbeispiel 9: Pop	49	Mehrkanal
	50	Downmix IRT_5ib
	51	Downmix IRT_5j
	52	Handmix
	53	Downmix ITU
	54	Downmix IRT_5keil
Hörbeispiel A: Intensitäts- stereofonie	55	Mehrkanal
	56	Downmix IRT_5ib
	57	Downmix IRT_5j
	58	Downmix ITU
	59	Downmix IRT_5keil
Hörbeispiel B: Laufzeit- stereofonie	60	Mehrkanal
	61	Downmix IRT_5ib
	62	Downmix IRT_5j
	63	Downmix ITU
	64	Downmix IRT_5keil
Hörbeispiel C: Kammfilter	65	Mehrkanal
	66	Downmix IRT_5ib
	67	Downmix IRT_5j
	68	Downmix ITU
	69	Downmix IRT_5keil

## Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mich in der Zeit dieser Diplomarbeit darin unterstützt haben, diese Arbeit zu realisieren.<sup>146</sup> Allen voran denke ich dabei an meine Betreuer Gerhard Stoll und Prof. Oliver Curdt.

Prof. Oliver Curdt, der mich das ganze Studium hindurch immer wieder ermutigt hat, mich weiter im Bereich der Tontechnik zu profilieren, mir im Rahmen von Vorlesungen, Seminaren und Studioproduktionen viele wertvolle Techniken und Tipps vermittelt und mein Gehör geschult hat. Gerhard Stoll, der mir in den letzten Monaten mit Rat und Tat zur Seite stand und mit seiner motivierenden Art immer wieder Mut, Kraft und neue Ideen in mir weckte.

Ganz besonders möchte ich mich bei Jörg Deigmöller bedanken der mich, als einer meiner Vorgänger am IRT zum Thema Downmix, mit hilfreichen Hintergrundinformationen und guten Ratschlägen versorgt hat, mich zu jedem Zeitpunkt unterstützt und mir neuen Mut gegeben hat, wenn der erhoffte Erfolg einmal ausgeblieben ist. Für das gute Ergebnis des IRT-Downmix ist auch ganz wesentlich Martin Keil verantwortlich, der mit seiner Diplomarbeit eine gewichtige Vorarbeit geleistet hat.

Für das konspirative Mitdenken bei technischen Problemen möchte ich mich bei Martin Link und Jens Groh bedanken, die mir in der für mich neuen Welt der digitalen Signalverarbeitung mit einem schier unerschöpflichen Wissen jederzeit zur Seite standen.

In diesem Zusammenhang möchte ich die schöne Zeit am Institut für Rundfunktechnik nicht unerwähnt lassen. Ganz wesentlich dazu beigetragen haben die freundlichen und kompetenten Mitarbeiter der Abteilung Audiosystemtechnik Gerhard Spikofski, Wolfgang Krafft, Klaus Laumann, Ralf Neudel ebenso wie meine ZimmergenossInnen Aylin Vogl, Michael Meyer und Jan-Eric Eichhorn. Und was wäre diese Zeit gewesen ohne die inspirierenden Gespräche mit Silvio Naumann und Hansi Huber.

Diese Arbeit hätte nicht so gigantische Formeln und ein ebenso schönes Titelblatt, gäbe es da nicht Matthias Schnepf, der mir als Mathematiker und Physiker zur Seite stand und Petra Riesemann, die mich in grafischen Angelegenheiten beraten hat.

Ich hätte wohl unter einer Isar-Brücke nächtigen müssen, gäbe es da nicht einen guten Freund, der mich während dieser Zeit herzlich in München aufgenommen hat. Manch lustiger Abend mit Matthias Strößner und Carmen Böhlke sowie Isabel, Franziska, Natalie und Ferdinand wird diese Zeit unvergessen machen.

Leider ist es heutzutage schwierig und nicht mehr selbstverständlich, ein Studium finanzieren zu können. Daher fühle ich mich in großem Maße gegenüber dem Studentenwerk Stuttgart zu Dank verpflichtet, das mir dieses Studium ermöglicht hat.

Schlussendlich möchte ich mich bei meinen Eltern Eveline und Günter Runow bedanken, die mir diese Ausbildung ermöglicht und mich in meinem gesamten Vorhaben treu begleitet haben. Eine herzliche und tiefe Verbundenheit möchte ich meiner Freundin Cornelia Egger aussprechen, die immer an mich geglaubt hat.

Vielen lieben Dank.

---

<sup>146</sup> Ich bitte die Weglassung von Dipl.-Ing., Dipl.-Tonmeister und ähnlichen Titeln aus Gründen der Übersichtlichkeit zu entschuldigen. ☺

