UNIVERSITÄT ZU LÜBECK

Aus dem Institut für Signalverarbeitung
der Universität zu Lübeck
Direktor: Prof. Dr.-Ing. Alfred Mertins

Neue Methoden zur kanalbasierten Raumhallkompensation

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck
Aus der Sektion Informatik / Technik

Vorgelegt von
Jan Ole Jungmann
aus Bad Oldesloe
Lübeck, 2014

1. Berichterstatter:	Prof. Dr.-Ing. Alfred Mertins
2. Berichterstatter:	Prof. Dr.-Ing. Erhardt Barth
Vorsitzender der Prüfungskommission:	Prof. Dr.-Ing. Andreas Schrader
Tag der mündlichen Prüfung:	17. Dezember 2014
Zum Druck genehmigt:	18. Dezember 2014

Zusammenfassung

Im Rahmen dieser Arbeit werden neue Verfahren zur qualitativ hochwertigen kanalbasierten Raumhallkompensation entwickelt. Des Weiteren wird ein Formelwerk zur gemeinsamen Raumhall- und Übersprechkompensation präsentiert.

Zum Entwurf der Filter wird üblicherweise ein auf der Zeitbereichsdarstellung der akustischen Kanäle basierendes Optimierungsproblem gelöst; eventuell auftretende spektrale Verzerrungen werden durch ein weiteres Filter kompensiert.

Im Rahmen dieser Arbeit wird gezeigt, dass die gleichzeitige Optimierung der Zeit- und Frequenzbereichsdarstellung der akustischen Kanäle einer getrennten Betrachtung von Zeit- und Frequenzbereichsdarstellung vorzuziehen ist. Es werden zwei neue Verfahren entwickelt, die eine Formung im Zeitbereich bei gleichzeitiger Optimierung des Betragsfrequenzgangs erlauben. Während die erste Methode auf der p-Norm basiert, optimiert das zweite Verfahren ein an das *Maß zur Bewertung der spektralen Flachheit* angelehntes Kriterium; hierdurch ist es möglich, beliebige Betragsfrequenzgänge vorzugeben.

Des Weiteren werden zwei neue Verfahren entwickelt, um den räumlichen Gültigkeitsbereich der entworfenen Filter zu vergrößern. Üblicherweise sind nach der Filteroptimierung schon kleinste Fehlplatzierungen ausreichend, um zu einem Gesamtsystem mit schlechteren Übertragungseigenschaften zu führen. Der Mehrpositions-Ansatz, der auf einer exzessiven Messung vieler Raumimpulsantworten innerhalb des Abhörvolumens basiert, ist in der Praxis üblicherweise nicht anwendbar. Die hier erarbeiteten Verfahren erlauben eine räumlich robuste Entzerrung auf Grundlage einer einzigen Messung des akustischen Kanals zu einer Referenzposition. Die erste Methode hierfür basiert auf einem stochastischen Fehlerterm, welcher in das Optimierungsproblem integriert wird. Für das zweite Verfahren werden auf Grundlage eines neu entwickelten Modells weitere Instanzen des gemessenen akustischen Kanals generiert und gemeinsam mit dem gemessenen Kanal optimiert.

Alle neu entwickelten Methoden werden mit gemessenen Raumimpulsantworten getestet; die erzielten Ergebnisse werden in dieser Arbeit präsentiert.

Abstract

The purpose of room impulse response reshaping is to reduce reverberation and thus to improve the perceived quality of the received signal by prefiltering the source signal before it is played with a loudspeaker. The filter design is usually carried out by solving a time-domain based optimization problem.

In this work, a framework is presented to jointly solve the task of listening room compensation and crosstalk cancellation. Therefore, a p-norm based objective function is proposed.

Furthermore, it is shown that for high-quality listening-room compensation it is neccessary to consider both the time- and the frequency-domain representations of the overall acoustic channels. Two new methods are developed to jointly optimize the time- and the frequency-domain representations of the overall systems. While one method is based on the p-norm, the second method is loosely based on the spectral flatness measure and allows to yield an overall acoustic channel with predefined shapes in the time and frequency domains.

Besides that, the problem of robustness against spatial misalignment is considered. Approaches to improve the robustness against small movements that are based on exhaustive measurements of room impulse responses yield good results, however, they are usually not feasible in practice. In this work two new methods are developed to improve the spatial robustness of reshaping filters. While the first approach is based on a stochastic modelling of the perturbations caused by microphone movement, the second method utilizes artifically perturbed instances of the acoustic channels.

Finally, results are given for all the developed methods in terms of dereverberation performance, *spectral flatness*, and spatial robustness.

Vorwort und Danksagung

Diese Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Signalverarbeitung an der Universität zu Lübeck. Ohne eine Reihe von Menschen wäre diese Arbeit so nicht realisierbar gewesen. Bei diesen Menschen möchte ich mich hier bedanken.

Als erstes möchte ich mich bei meinem Doktorvater Herrn Prof. Dr.-Ing. Alfred Mertins bedanken, der mir die Möglichkeit gegeben hat, die Forschungen für diese Arbeit durchzuführen. Er stand mir hierbei immer mit Rat und Tat zur Seite und hatte manchmal mehr Vertrauen in meine Arbeit als ich selbst.

Bei meinen jetzigen und ehemaligen Arbeitskollegen möchte ich mich für die nette Aufnahme bedanken. Zu vielen Arbeitskollegen ist ein freundschaftliches Verhältnis entstanden, welches hoffentlich noch lange anhält. Besonders hervorheben möchte ich an dieser Stelle Radoslaw Mazur und Florian Müller. Durch seine motivierenden Worte und viele fachliche Diskussionen hat Radek nicht unwesentlich dazu beigetragen, dass diese Arbeit fertig gestellt wurde. Bei Florian möchte ich mich für das gute *Zusammenleben* im Büro bedanken. Mein Dank gilt natürlich auch meinen anderen Arbeitskollegen. Namentlich erwähnen möchte ich hier Lei Chen, Olaf Christ, Alexandru Condurache, Dierck Matern, Marco Maaß, Huy Phan, Anita Möller und Christiane Ehlers. Bei Martin Nopper möchte ich mich für die Unterstützung im Rahmen seiner Tätigkeit als studentische Hilfskraft bedanken.

Natürlich gebührt ein großer Dank meiner Frau Claudia sowie unserem gemeinsamen Sohn Jonas – beide haben mein Leben auf vielfältige Art und Weise bereichert. Insbesondere bei Claudia möchte ich mich dafür bedanken, dass sie mir die Zeit gegeben hat, diese Arbeit zu verfassen.

Ein weiterer Dank gilt meinen Eltern, die mich in allen meinen Entscheidungen bekräftigt haben und immer hinter mir standen. Ohne euch wäre ich wahrscheinlich nicht diesen Weg gegangen!

Danke!

Inhaltsverzeichnis

1 Einleitung **1**

1.1 Motivation 1

1.2 Abgrenzung der Arbeit 3

1.3 Aufbau der Arbeit 4

1.4 Notation 5

2 Grundlagen **9**

2.1 Übertragung akustischer Signale 9

2.1.1 Freifeldmodell 9

2.1.2 Schallausbreitung in geschlossenen Räumen 10

2.1.3 Raumimpulsantwort 12

2.1.3.1 Simulation von Raumimpulsantworten 14

2.1.3.2 Messung von Raumimpulsantworten 14

2.2 Temporale Maskierungseigenschaften des menschlichen Hörapparates . 20

2.3 Bewertung akustischer Übertragungskanäle 22

2.3.1 Maße zur Bewertung der Zeitbereichsdarstellung 22

2.3.2 Bewertung der Frequenzbereichsdarstellung 26

2.4 Zusammenfassung 28

3 Kanalbasierte Raumhallkompensation **29**

3.1 Übersicht 29

3.2 Kanalkompensation mit gegebenem Zielsystem 31

3.2.1 Methode des kleinsten quadratischen Fehlers 32

3.2.1.1 Herleitung 32

3.2.1.2 Optimierung des Systemabstandes 35

3.2.1.3 Optimierung des gewichteten Systemabstandes 37

3.2.2 Optimierung der Maximumsnorm 39

3.2.3 Zusammenfassung 42

3.3 Kanalkompensation ohne Vorgabe eines Zielsystems 42
3.3.1 Quadratisches Optimalitätskriterium 43
3.3.2 Spektrale Aspekte 46
3.3.3 Maximumsnormbasierte Formung 47
3.3.3.1 Impulsantwortformung mit einer maximumsnormbasierten Nebenbedingung 48
3.3.3.2 Formung durch Optimierung der Maximumsnorm . . 50
3.3.4 Formung durch Optimierung der p-Norm 53
3.3.4.1 Gewichtsfenster 54
3.3.5 Zusammenfassung 56
3.4 Robuste Kanalkompensation 58
3.4.1 Beschreibung von fehlerbehafteten Kanälen 59
3.4.2 Robustheit gegen Kanalfehlschätzungen 60
3.4.3 Räumliche Robustheit 62
3.4.4 Untersuchungen zur räumlichen Robustheit 63
3.4.5 Erhöhung der Robustheit gegenüber Fehlplatzierung durch multiple Messungen 66
3.4.6 Weitere Ansätze zur Erhöhung der Robustheit gegenüber Fehlplatzierung 70
3.4.7 Zusammenfassung 72
3.5 Übersprechkompensation 73
3.5.1 Quadratisches Optimalitätskriterium für den Entwurf der Filter zur Übersprechkompensation 75
3.5.2 Weitere Ansätze zur Übersprechkompensation 77
3.6 Zusammenfassung 78

4 Erweiterte Methoden 79
4.1 Mehrkanalige Formung der Impulsantworten im Zeitbereich 79
4.1.1 Formulierung im Zeitbereich 82
4.1.2 Entwurf der Filter für die Formung im Zeitbereich 84
4.1.2.1 Kostenfunktion für die Formung im Zeitbereich . . . 85
4.1.2.2 Herleitung der Gradienten für die Formung im Zeitbereich 86

4.2 Regularisierung im Spektralbereich 88
4.2.1 Regularisierung auf Basis der p-Norm 90
4.2.1.1 Herleitung des Gradienten 91
4.2.1.2 Erweiterung der auf der p-Norm basierenden Regularisierung auf beliebig viele Kanäle 93
4.2.2 Spektrale Flachheit als Kriterium 95
4.2.2.1 Generalisiertes Maß zur Bewertung der spektralen Flachheit 95
4.2.2.2 Regularisierung durch das generalisierte SFM 97
4.2.2.3 Gradient für das generalisierte SFM 97
4.2.3 Zusammenfassung 100
4.3 Räumliche Robustheit 101
4.3.1 Robustheit durch Fehlermodellierung 102
4.3.1.1 Fehlermodell nach Radlović et al. 103
4.3.1.2 Herleitung des Fehlerterms 104
4.3.1.3 Integration des Fehlerterms in den Filterentwurf . . . 107
4.3.1.4 Erweiterung auf beliebig viele Kanäle 109
4.3.1.5 Effiziente Implementierung 110
4.3.2 Perturbationsbasierter Ansatz 111
4.3.2.1 Beschreibung der Perturbationen 111
4.3.2.2 Modellierung der Perturbation im Zeitbereich 112
4.3.2.3 Modellierung der Perturbation im Frequenzbereich . . 113
4.3.2.4 Erzeugung perturbierter Raumimpulsantworten . . . 114
4.4 Wahl der Gewichtsfenster 116
4.4.1 Übersprechkompensation 118
4.5 Zusammenfassung 120

5 Daten 121
5.1 Verwendete Komponenten 121
5.2 Messaufbau 121
5.3 Durchführung 122
5.4 Weitere Raumimpulsantworten 125

6 Experimente und Ergebnisse 129
6.1 Neue Maße zur objektiven Bewertung der Ergebnisse 129
6.1.1 Quantifizierung des wahrnehmbaren Nachhalls 129

6.1.2 Bewertung der Übersprechkompensation 131
6.2 Ergebnisse 132
6.2.1 Optimierung der Frequenzdarstellung 132
6.2.1.1 Regularisierung auf Basis der p-Norm 132
6.2.1.2 Regularisierung auf Basis der spektralen Flachheit . . 139
6.2.2 Räumliche Robustheit 146
6.2.2.1 Räumliche Robustheit durch Abtastung des Hörvolumens 147
6.2.2.2 Räumliche Robustheit durch statistische Modellierung 149
6.2.2.3 Räumliche Robustheit durch perturbierte Raumimpulsantworten 153
6.2.3 Übersprechkompensation 156
6.3 Zusammenfassung 161

7 Diskussion und Fazit 163
7.1 Diskussion 163
7.2 Zusammenfassung und Fazit 166

8 Ausblick 169

Literatur 171

1 Einleitung

In diesem Kapitel wird eine Übersicht über den Inhalt dieser Arbeit gegeben. Hierzu wird die Kompensation von Übertragungssystemen im Allgemeinen kurz umrissen und es wird im Speziellen die Kompensation akustischer Kanäle motiviert. Da es neben den im Rahmen dieser Arbeit entwickelten Verfahren weitere Klassen von Algorithmen zur Kompensation akustischer Kanäle gibt, wird in Abschnitt 1.2 eine Abgrenzung dieser Arbeit gegenüber anderen Arbeiten und Verfahren gegeben. Des Weiteren wird in Abschnitt 1.3 der Aufbau dieser Dissertation vorgestellt; die verwendete Notation wird in Abschnitt 1.4 eingeführt.

1.1 Motivation

Kommunikation und damit die Übermittlung von Informationen und Daten ist ein integraler Bestandteil der Welt. Jeder Mensch ist auf einen funktionierenden Informationsaustausch innerhalb seines Körpers angewiesen; ist beispielsweise die fehlerfreie Übermittlung von Befehlen aus dem Gehirn an die einzelnen Organe und Muskeln nicht gewährleistet, so ist mit massiven Einschränkungen der Lebensqualität zu rechnen. Einen vergleichbar gewichtigen Wert nimmt die zwischenmenschliche Kommunikation ein. Ohne einen Erfahrungs- und Wissensaustausch hätte sich die Gesellschaft nicht zu ihrer heutigen Form entwickelt.

Die zwischenmenschliche Kommunikation erfolgte lange Zeit lediglich mündlich - also durch Sprache. Die gesprochene Sprache ist den textbasierten Kommunikationsformen (Brief, eMail, Chat, SMS, ...) bezüglich des Informationsgehaltes weit überlegen. Emotionen und Gemütszustände lassen sich sehr differenziert über die Sprachmelodie und Betonung übermitteln. Das geschriebene Wort kann diese *Metainformationen* nicht adäquat abbilden und transportieren. An der Kommunikation über menschliche Sprache sind drei Komponenten beteiligt: (1) ein Sprecher der Schallwellen aussendet; (2) ein Medium, welches die Schallwellen transportiert und (3) ein Zuhörer, der die Schallwellen wahrnimmt. Die Beschaffenheit des Übertragungsmediums bestimmt den Höreindruck des Zuhörers und kann zu massiven Einschränkungen für das Sprachverständnis führen.

So ist beispielsweise aufgrund von Nachhall ein Sprecher in einer großen leeren Halle bei weitem schlechter zu verstehen als in einem kleinen Büroraum. Der Nachhall wird dem Sprachsignal durch die Übertragung in einem geschlossenen Raum aufgrund mehrfacher Reflexionen der Schallwellen an den Wänden sowie an Einrichtungsgegenständen aufgeprägt.

Gegenstand dieser Arbeit ist es, den Einfluss des Übertragungskanals durch den Entwurf geeigneter Filter zu kompensieren. Als potentielle Anwendungsgebiete der in dieser Arbeit entwickelten Verfahren sind Freisprecheinrichtungen und Audiosysteme für den Unterhaltungsbereich zu nennen, die einen qualitativ hochwertigen und räumlichen Höreindruck vermitteln und gleichzeitig, im Falle von Sprachsignalen, eine hohe Sprachverständlichkeit garantieren sollen.

Reden mehrere Sprecher gleichzeitig, so kommt es neben dem Nachhall noch zu sogenannten Übersprecheffekten – der Zuhörer nimmt eine Überlagerung der einzelnen Sprecher wahr, was die Verständigung weiter erschwert. Die Methoden zur kanalbasierten Raumhallkompensation lassen sich üblicherweise auch auf das Problem der Übersprechkompensation anwenden.

Auch in vielen anderen Anwendungsbereichen kann die Messung von Daten ähnlich beschrieben werden, nämlich durch eine Signalquelle, der Übertragung des Signals durch einen nicht perfekten Übertragungskanal und einen Empfänger. In der Astronomie beispielsweise wird das von weit entfernten Sternen ausgesendete Licht durch Staubpartikel gestreut. Ein Beobachter auf der Erde wird somit immer ein gestörtes Bild wahrnehmen. Zur Verbesserung der Qualität muss der Einfluss des Übertragungskanals kompensiert werden. Auch in der elektronischen Datenübertragung über (Funk-) Netzwerke im Allgemeinen ist die Kompensation der Einflüsse der Übertragungskanäle von großer Bedeutung. Ein nicht perfekter Übertragungskanal führt zu einer verminderten Übertragungsrate; Anwendungen wie hochauflösendes digitales Fernsehen sind auf eine hohe Datenrate sowie eine möglichst fehlerfreie Datenübertragung angewiesen.

Ein weiteres essentielles Anwendungsgebiet der Kanalkompensation ist die Medizin. Bei nahezu jeder nichtinvasiven Untersuchung des menschlichen Körpers werden Signale aus dem Körperinneren mit Geräten außerhalb des Körpers aufgenommen. So werden beispielsweise die vom Herz oder der Lunge erzeugten Töne mit einem Stethoskop an der Körperoberfläche abgegriffen. Bei der Elektroenzephalografie werden elektrische Signale des Gehirns mit Elektroden an der Oberfläche des Kopfes aufgenommen. Neben der Degradation durch die Übertragung durch die Schädeldecke kommt es auch hier zu

Übersprecheffekten – jede Elektrode misst eine Überlagerung der Signale von mehreren Hirnarealen.

Die Kompensation von Übertragungskanälen im Allgemeinen ist somit von großer praktischer Bedeutung und Gegenstand aktueller Forschungsbemühungen.

1.2 Abgrenzung der Arbeit

Die im Rahmen dieser Arbeit entwickelten Verfahren basieren auf der Annahme, dass alle beteiligten Systeme linear sind und sich hinreichend genau mit einer endlich langen Impulsantwort (FIR) beschreiben lassen. Zum Entwurf der Filter ist eine vorherige Messung oder Schätzung der Übertragungskanäle nötig. Die entworfenen Filter können in einem Vorfilternetzwerk vor die Lautsprecher geschaltet werden, um somit dem Zuhörer einen qualitativ hochwertigen Höreindruck zu vermitteln. Als mögliche Anwendungsgebiete sind beispielsweise das Heimkino oder Telefonkonferenzsysteme mit Freisprecheinrichtungen zu nennen. Werden die entworfenen Filter zur Nachbearbeitung eines aufgenommenen Signals verwendet, kann beispielsweise die Erkennungsrate eines Spracherkenners gesteigert werden [27, 41].

Wird explizit von einer Nachbearbeitung aufgenommener Signale ausgegangen, so eröffnen sich neue Möglichkeiten zur Raumhallkompensation. Liegen mehrere Messungen eines einzigen Quellensignals vor, so können die einzelnen Übertragungskanäle *blind*, also ohne Kenntnis des Quellensignals, aus den Aufnahmen geschätzt werden [54, 56, 57, 65]. Auf Basis der blind geschätzten Kanalimpulsantworten können dann die Filter zur Kanalkompensation entworfen werden. Ist weiterhin bekannt, dass es sich bei dem Quellensignal um ein Sprachsignal handelt, so können spezielle Methoden zur blinden Enthallung von Sprachsignalen angewendet werden [26, 69, 129]. Die hier entwickelten Verfahren sollen allerdings allgemeingültig für jede Art von Quellensignalen anwendbar sein – nicht nur für Sprachsignale.

Unabhängig des Quellensignals eignen sich nichtlineare Verfahren wie die *spektrale Subtraktion* lediglich zur Enthallung von aufgenommenen Audiosignalen [36, 52, 53, 92, 99]. Die spektrale Subtraktion basiert auf der Manipulation des Kurzzeit-Betragsspektrums des aufgenommenen Signals. Algorithmen dieser Art eignen sich nicht für die Prozessierung der Quellensignale vor der Lautsprecherreproduktion und sind damit nicht Gegenstand dieser Arbeit.

Die Verwendung als Vorfilter erzwingt eine vorherige Identifikation der akustischen Übertragungskanäle. Somit sind die hier entwickelten Methoden per-se nicht *blind*.

Im Zusammenhang mit einer blinden Kanalschätzung können die hier entwickelten Verfahren jedoch auch *blind* verwendet werden.

Weitere Ansätze zur Kontrolle eines Schallfeldes sind die sogenannte *Wellenfeldsynthese* [8, 10, 153, 155] sowie die *Ambisonics höherer Ordnung* [22, 23]; ein Vergleich dieser Methoden wird beispielsweise in [154, 156] gegeben. Bei diesen Methoden wird ein Volumen von einer Vielzahl von Lautsprechern eingeschlossen. Aufgrund der praktischen Nutzbarkeit ist auch diese Klasse von Verfahren nicht Gegenstand dieser Arbeit; der Anspruch an die hier entwickelten Methoden ist, schon mit einer moderaten Anzahl an Lautsprechern ein hochwertiges Hörerlebnis zu ermöglichen.

1.3 Aufbau der Arbeit

Nachdem in Kapitel 1 die generelle Problemstellung der Kompensation von Übertragungskanälen und im Speziellen die Kompensation von akustischen Übertragungskanälen motiviert wurde, wird in Kapitel 2 auf die Entstehung von Nachhall und die Beschreibung von akustischen Kanälen eingegangen. Hierzu wird der Begriff der *Raumimpulsantwort* erläutert. Des Weiteren werden Methoden zur Messung der Raumimpulsantwort und quantitative Maße zur Bewertung akustischer Kanäle vorgestellt.

In Kapitel 3 werden Methoden zur kanalbasierten Raumhallkompensation präsentiert. Dieses Kapitel gibt einen Überblick über verschiedene Ansätze und Algorithmen zum Entwurf der Kompensationsfilter. Die mit der Kanalkompensation eng verwandte Problemstellung der *Übersprechkompensation* wird in Kapitel 3.5 behandelt.

Den Kern dieser Dissertation bildet Kapitel 4. In diesem Kapitel werden die neuen Verfahren zur robusten, qualitativ hochwertigen Raumhallkompensation und Übersprechkompensation entwickelt. Hierzu wird ein einheitliches mathematisches Rahmenwerk entwickelt. Die neu entwickelten Verfahren basieren zum Teil auf den in den Kapiteln 2 und 3 vorgestellten Grundlagen und Methoden.

Die mit den neu entwickelten Verfahren durchgeführten Experimente sowie die erzielten Ergebnisse werden in Kapitel 6 beschrieben; in diesem Kapitel wird auch ein neu entwickeltes Maß zur quantitativen Bewertung des wahrnehmbaren Nachhalls vorgestellt und zur Bewertung der Ergebnisse herangezogen. Die für die Experimente benötigten Daten werden in Kapitel 5 beschrieben. Den Abschluss dieser Arbeit bildet eine Diskussion und vergleichende Bewertung der neu entwickelten Verfahren in Kapitel 7 sowie ein Ausblick für mögliche weiterführende Arbeiten in Kapitel 8.

1.4 Notation

Innerhalb dieser Arbeit wird eine konsistente Notation verwendet. Sofern nicht explizit anders angegeben, gelten hierbei folgende Regeln:

- Skalare werden mit (lateinischen oder griechischen) kursiven Klein- oder Großbuchstaben bezeichnet. Beispiel: b, A, α.
- Signale und Filter im Zeitbereich werden durch kursive Kleinbuchstaben bezeichnet. Der Buchstabe n wird üblicherweise als diskreter, der Buchstabe t als kontinuierlicher Zeitindex verwendet; sofern nicht anders angegeben sind alle in dieser Arbeit verwendeten Signale und Filter zeitdiskret. Beispiele: $x(n)$, $s(t)$.
- Die Länge eines Filters im Zeitbereich wird mit einem großen L mit entsprechendem Subskript bezeichnet. Beispiel: L_c bezeichnet die Länge des Filters $c(n)$.
- Vektoren werden mit fetten Kleinbuchstaben bezeichnet. Beispiel: $\boldsymbol{x}$.
- Matrizen werden mit fetten Großbuchstaben bezeichnet. Beispiel: $\boldsymbol{C}$.
- Einzelne Komponenten eines Vektors oder einer Matrix werden durch Kleinbuchstaben mit Subindices bezeichnet. Beispiel: x_5 bezeichnet das fünfte Element in dem Vektor $\boldsymbol{x}$. Mit c_{mn} wird das Element in der m-ten Zeile und n-ten Spalte der Matrix $\boldsymbol{C}$ bezeichnet.
- Der Nullvektor (beziehungsweise je nach Kontext die Nullmatrix) wird als $\mathbf{0}$ bezeichnet.
- Die Einheitsmatrix wird als $\boldsymbol{I}$ bezeichnet.
- Die diskrete Fourier-Transformierte eines diskreten Signals oder Filters $c(n)$ wird als $C(e^{j\omega_k})$ bezeichnet, wobei ω_k eine diskrete Frequenz ist; mit k als diskretem Frequenzindex gilt kurz $C(k) = C(e^{j\omega_k})$.
- Werden Vektoren oder Matrizen konkateniert, so werden sie in einem größeren Vektor (beziehungsweise einer Matrix) angeordnet. Beispielsweise definiert

$$\boldsymbol{h} = \begin{bmatrix} \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \end{bmatrix} \tag{1.1}$$

einen Vektor $\boldsymbol{h}$, der erst die Elemente von $\boldsymbol{h}_1$ und darunter die Elemente von $\boldsymbol{h}_2$ enthält.

- Die Transponierte sowie die hermitesch Transponierte einer Matrix $\boldsymbol{A}$ wird mit $\boldsymbol{A}^T$, beziehungsweise $\boldsymbol{A}^H$ bezeichnet.

- Die Inverse einer Matrix $\boldsymbol{A}$ wird als $\boldsymbol{A}^{-1}$ bezeichnet.

- Die Moore-Penrose Pseudoinverse einer Matrix $\boldsymbol{A}$ wird als $\boldsymbol{A}^+$ notiert und ist im hier relevanten überbestimmten Fall durch $\boldsymbol{A}^+ = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T$ gegeben [17, 134].

- Der Erwartungswert wird mit dem Operator $\mathrm{E}\{\cdot\}$ bezeichnet.

- Mit $\|\boldsymbol{x}\|_p$ wird die p-Norm (auch ℓ_p-Norm) eines Vektors bezeichnet. Die p-Norm eines Vektors $\boldsymbol{x} = [x_1, \ldots, x_N]^T$ berechnet sich als

$$\|\boldsymbol{x}\|_p = \left(\sum_{n=1}^{N} |x_n|^p\right)^{\frac{1}{p}}. \tag{1.2}$$

- Die Maximums- oder Unendlich-Norm eines Vektors $\boldsymbol{x} = [x_1, \ldots, x_N]^T$ wird mit $\|\boldsymbol{x}\|_\infty$ notiert und ist als

$$\|\boldsymbol{x}\|_\infty := \max_n \; |x_n| \tag{1.3}$$

  definiert.

- Die Null-Pseudonorm (auch ℓ_0-Pseudonorm) eines Vektors $\boldsymbol{x} = [x_1, \ldots, x_N]^T$ wird als $\|\boldsymbol{x}\|_0$ notiert und gibt die Anzahl der Elemente $x_n \neq 0$ zurück.

- Der Operator $\mathrm{diag}\{\cdot\}$ transformiert einen Vektor in eine Diagonalmatrix. Für einen Vektor $\boldsymbol{x}$ der Länge N gilt also

$$\mathrm{diag}\{\boldsymbol{x}\} = \begin{bmatrix} x_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x_N \end{bmatrix}. \tag{1.4}$$

Wird $\mathrm{diag}\{\cdot\}$ auf eine quadratische Matrix $\boldsymbol{X}$ der Dimension $N \times N$ angewendet, so wird ein Vektor $\boldsymbol{x}$ der Länge N zurückgegeben, der die Elemente auf der Hauptdiagonalen der Matrix $\boldsymbol{X}$ enthält:

$$\mathrm{diag}\{\boldsymbol{X}\} = [x_{11}, \ldots, x_{NN}]^T \,. \tag{1.5}$$

- Der Operator $\mathrm{sign}\{\cdot\}$ liefert die Vorzeichen des Arguments. Wird ein Vektor als Argument übergeben, so liefert $\mathrm{sign}\{\cdot\}$ einen Vektor mit den Vorzeichen der einzelnen Einträge. Wird der Operator $\mathrm{sign}\{\cdot\}$ auf eine komplexe Zahl angewendet, so gibt er das komplexe Vorzeichen zurück. Das komplexe Vorzeichen einer Zahl ist die Projektion der Zahl auf den Einheitskreis.

- Der relle und der imaginäre Teil einer komplexen Zahl $z \in \mathbb{C}$, $z = a + jb$ wird durch $\Re\{\cdot\}$ und $\Im\{\cdot\}$ zurück gegeben. Es gilt

$$\Re\{z\} = a \quad \text{beziehungsweise} \quad \Im\{z\} = b. \tag{1.6}$$

- Das Kronecker-Delta wird mit δ_{ij} notiert und ist als

$$\delta_{ij} = \begin{cases} 1, & \text{falls } i = j \\ 0, & \text{falls } i \neq j \end{cases} \tag{1.7}$$

definiert.

2 Grundlagen

In diesem Kapitel wird die Entstehung von Nachhall beschrieben. Grundlage hierfür bildet die Schallausbreitung in geschlossenen Räumen. Der akustische Kanal von einer Schallquelle zu einem Empfänger beziehungsweise Zuhörer wird als lineares zeitinvariantes System durch die Raumimpulsantwort beschrieben.

In einem sehr einfachen Szenario, dem sogenannten Schuhkarton-Modell (englisch *shoebox model*), können Raumimpulsantworten effizient simuliert werden [2]. Methoden zur Simulation sowie zur Messung von Raumimpulsantworten werden in Abschnitt 2.1.3.1 beziehungsweise in Abschnitt 2.1.3.2 vorgestellt.

Des Weiteren werden in diesem Kapitel einige der in der Literatur üblicherweise verwendeten Maße zur Bewertung und Charakterisierung von akustischen Übertragungskanälen vorgestellt.

2.1 Übertragung akustischer Signale

Wird ein akustisches Signal von einer Quelle zu einem Empfänger übertragen, so wird allgemein zwischen Freifeldbedingungen und der Schallübertragung in geschlossenen Räumen unterschieden. Beide Szenarien werden kurz in den folgenden Unterabschnitten eingeführt.

Der Einfachheit halber werden im Rahmen dieser Arbeit die Übertragungseigenschaften der verwendeten Lautsprecher und Mikrofone nicht explizit betrachtet. Sofern die Lautsprecher und Mikrofone als lineare Systeme angenommen werden, können ihre Eigenschaften allerdings in der Raumimpulsantwort (siehe Abschnitt 2.1.3) zusammengefasst werden.

2.1.1 Freifeldmodell

Das Freifeldmodell beschreibt die Ausbreitung von Schallwellen in einer absolut *freien* Umgebung. Es wird idealisierend von einer punktförmigen Quelle und einem punktför-

migen Empfänger ausgegangen. In Abbildung 2.1 ist das Freifeldmodell exemplarisch dargestellt.

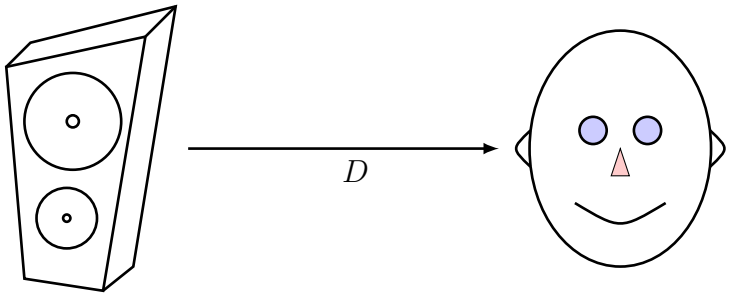

Abbildung 2.1: Skizze des Freifeldmodells. Der Schall erreicht ungehindert den Empfänger, der sich im Abstand D von der Schallquelle befindet.

Unter Freifeldbedingungen misst der Empfänger lediglich eine verzögerte und skalierte Version des Quellensignals. Die Verzögerung ist abhängig von der Schallgeschwindigkeit v und dem Abstand D zwischen Sender und Empfänger. Die Skalierung ist lediglich von dem Abstand D abhängig, wobei die Intensität mit dem Quadrat der Entfernung abnimmt. Wird das Quellensignal als $s(t)$ bezeichnet, so lässt sich das vom Empfänger gemessene Signal als

$$x(t) = D^{-2}s(t - D/v) \tag{2.1}$$

beschreiben.

2.1.2 Schallausbreitung in geschlossenen Räumen

Wird die Schallausbreitung in geschlossenen Räumen beschrieben, so müssen neben der Direktschallkomponente, also dem Schallanteil, der den Empfänger auf direktem Wege erreicht, noch die Reflexionen der Schallwellen an den Wänden und Gegenständen innerhalb des Raumes beachtet werden. In Abbildung 2.2 ist exemplarisch die Schallausbreitung in einem geschlossenen Raum dargestellt.

Wie unter Freifeldbedingungen (siehe Abbildung 2.1) erreicht den Empfänger als erstes die Direktschallkomponente (grün). Darüber hinaus werden die Schallwellen an den Wänden des Raumes reflektiert, bevor sie den Empfänger erreichen; die an lediglich einer Wand reflektierten Schallwellen (blau) werden als *frühe* Reflexionen bezeichnet. Werden die Schallwellen an mehr als einer Wand reflektiert (rot), so werden diese als *späte* Reflexionen bezeichnet. Die von den Schallwellen zurückgelegte Strecke bestimmt, wie auch bei der Direktschallkomponente, die Verzögerung und die Skalierung des von

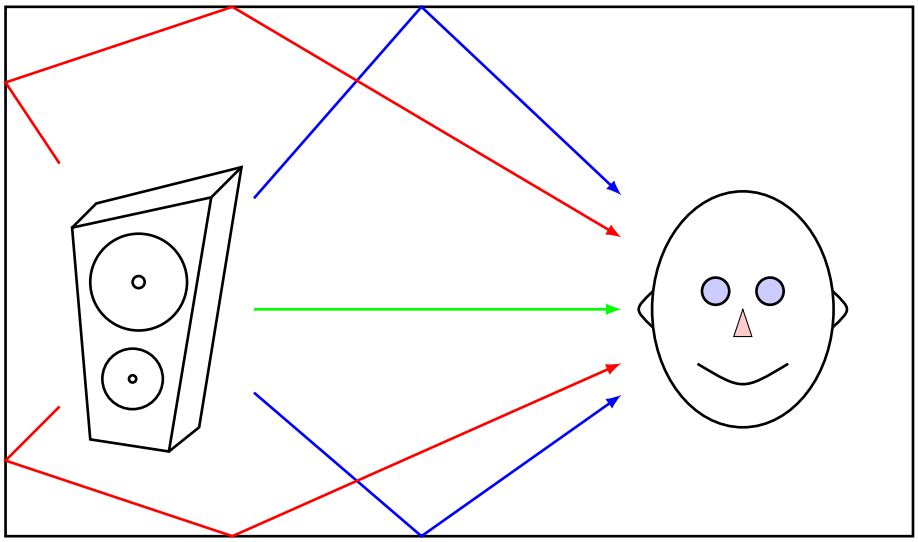

Abbildung 2.2: Skizze des generellen Aufbaus der in dieser Arbeit betrachteten akustischen Systeme; hier exemplarisch mit einer Schallquelle (Lautsprecher) und einer Abhörposition in einem geschlossenen Raum. Grün: Direktschallkomponente, blau: frühe Reflexionen, rot: späte Reflexionen.

dem Empfänger gemessenen Signals [25, 71, 124, 158]. Bei jeder Reflexion verlieren die Schallwellen zusätzlich Energie; der Energieverlust ist hierbei abhängig von den Reflexionseigenschaften der Wände.

Die absolute Verzögerung sowie die betragsmäßige Skalierung der Reflexionen können in einem Reflektogramm aufgetragen werden. Abbildung 2.3 zeigt ein typisches Reflektogramm. Der späte Nachhall (rot in Abbildung 2.3) ist die Hauptursache für eine verminderte Sprachverständlichkeit [139].

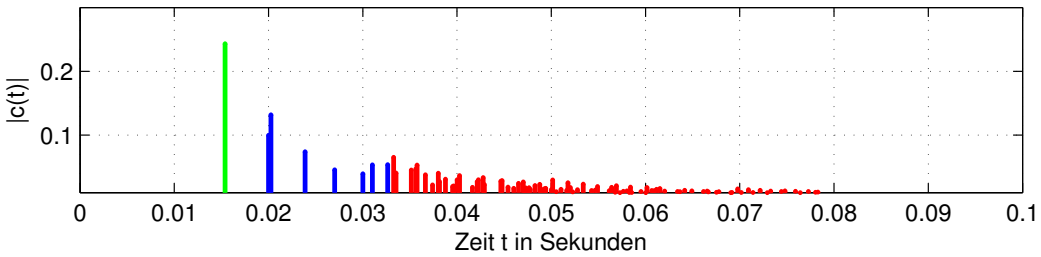

Abbildung 2.3: Darstellung eines Reflektogramms bestehend aus der Direktschallkomponente (grün), den frühen Reflexionen (blau) und den späten Reflexionen (rot).

Der Empfänger misst letztendlich eine Superposition von verzögerten und skalierten Versionen des Quellensignals. Der Übertragungskanal kann dementsprechend durch eine Impulsantwort, die Raumimpulsantwort, beschrieben werden.

Eine exakte Definition für die Differenzierung in frühe und späte Reflexionen gibt es nicht. In der Literatur werden üblicherweise die Reflexionen, die innerhalb der ersten 80 Millisekunden nach dem Direktschall beim Empfänger eintreffen, als die frühen Reflexionen bezeichnet; alle folgenden Reflexionen werden als späte Reflexionen oder diffuser Nachhall bezeichnet [60]. Das Formulieren von alternativen Vorschriften zur Einteilung der Raumimpulsantwort in frühe und späte Reflexionen ist immer noch Gegenstand aktueller Forschungsbemühungen [60].

2.1.3 Raumimpulsantwort

Eine Raumimpulsantwort beschreibt die Eigenschaften eines akustischen Kanals innerhalb eines Raumes. Der akustische Kanal wird durch die Position der Schallquelle und des Empfänger sowie durch die akustischen Eigenschaften des Raumes definiert; auch die Temperatur hat einen Einfluss auf die Raumimpulsantwort. Der Einfluss einer Änderung der Temperatur auf die Raumimpulsantwort wird im Rahmen dieser Arbeit nicht betrachtet. Der typische Aufbau eines Lautsprecher-Raum-Mikrofon (LRM) Systems ist in Abbildung 2.4 dargestellt. Das Quellensignal $s(n)$ wird über einen Lautsprecher wiedergegeben, über den akustischen Kanal $c(n)$, die Raumimpulsantwort, übertragen und von einem Mikrofon aufgenommen. Das in der Praxis immer vorhandene Rauschen $\eta(n)$ wird im Rahmen dieser Arbeit nicht betrachtet.

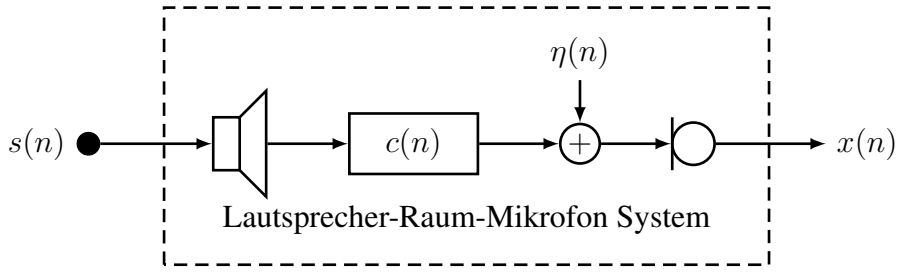

Abbildung 2.4: Darstellung eines einkanaligen Lautsprecher-Raum-Mikrofon Systems.

Das vom Empfänger gemessene Signal $x(n)$ wird durch das Faltungsprodukt des Quellensignals mit der (kausalen) Raumimpulsantwort beschrieben:

$$x(n) = \sum_{m=0}^{\infty} c(m)\, s(n-m) = s(n) * c(n)\,. \tag{2.2}$$

In Abbildung 2.5 ist eine gemessene Raumimpulsantwort in verschiedenen Darstellungsformen abgebildet. Für die visuelle Bewertung einer Raumimpulsantwort sind die logarithmische Darstellung der quadrierten Zeitkoeffizienten (Abbildung 2.5, mitte) und die logarithmische Darstellung des quadrierten Betragsfrequenzgangs (Abbildung 2.5, unten) von Bedeutung.

In der Abbildung 2.5 (mitte) ist das Abklingverhalten, also die Abnahme der Energie über die Zeit, der Raumimpulsantwort gut zu erkennen. In logarithmischer Darstellung ist das Abklingverhalten nahezu linear. Eine Raumimpulsantwort ist theoretisch unendlich lang. Für eine handhabbare Beschreibung ist es notwendig, die Raumimpulsantwort auf eine endliche Anzahl an Koeffizienten zu begrenzen.

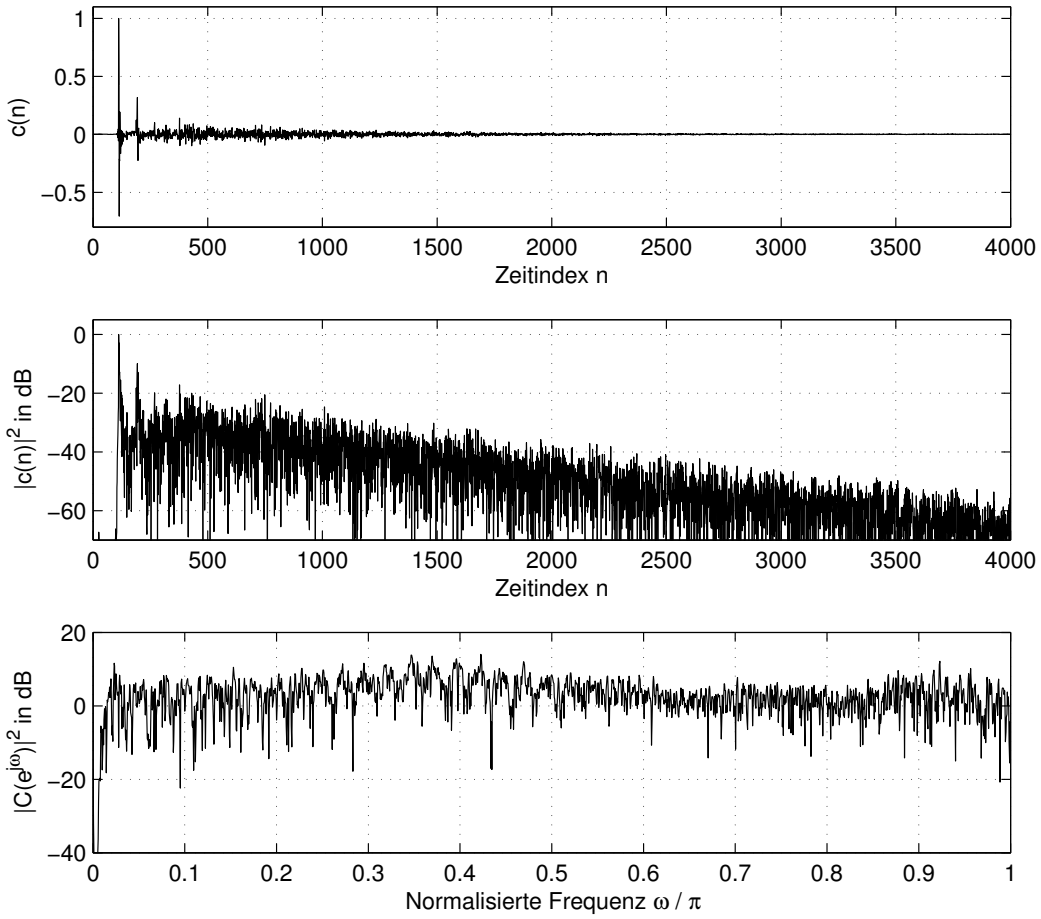

Abbildung 2.5: Verschiedene Darstellungsformen einer Raumimpulsantwort: (Oben) Darstellung auf einer linearen Skala, (Mitte) Darstellung der Energie der Zeitkoeffizienten auf einer logarithmischen Skala, (Unten) Darstellung des quadrierten Betragsfrequenzgangs auf logarithmischer Skala.

Raumimpulsantworten können dazu verwendet werden, die akustischen Eigenschaften eines Raumes zu untersuchen (siehe Abschnitt 2.3). In der virtuellen Akustik werden gemessene oder simulierte Raumimpulsantworten dazu verwendet, um virtuelle akustische Umgebungen glaubwürdig zu simulieren.

2.1.3.1 Simulation von Raumimpulsantworten

Zur Simulation von Raumimpulsantworten wird oftmals die Spiegelquellenmethode nach Allen und Berkley [2] verwendet. Allen und Berkley haben einen Algorithmus zur effizienten Simulation von Raumimpulsantworten in quaderförmigen Räumen entwickelt; das angenommene Modell wird auch als *shoebox-model* bezeichnet. Abbildung 2.6 stellt die Idee hinter dem Algorithmus dar.

Der Name des Modells stammt von seiner Funktionsweise: Der Raum wird mehrfach in alle Raumdimensionen *gespiegelt*. Jede Reflexion kann dann einfach durch die Berechnung des euklidischen Abstandes zwischen der virtuellen Quelle und dem Empfänger berechnet werden. Die Anzahl an Reflexionen an den Wänden entspricht der Ordnung der Spiegelung. Somit kann auch die durch die Reflexion an den Wänden verursachte Abschwächung des Signals in die Berechnung der Raumimpulsantwort einbezogen werden.

Die Spiegelquellenmethode ermöglicht lediglich die Simulation von Raumimpulsantworten innerhalb eines quaderförmigen Raumes. In Abbildung 2.7 ist eine simulierte Raumimpulsantwort auf einer logarithmischen Skala dargestellt. Deutlich zu erkennen ist der (auf einer logarithmischen Skala) lineare Abfall der Energie über die Zeit. Zur Simulation der Raumimpulsantwort wurde die Software von Emanuël Habets[1] verwendet, welche die Spiegelquellenmethode nach [2] implementiert.

2.1.3.2 Messung von Raumimpulsantworten

Von praktischer Bedeutung ist die Messung von Raumimpulsantworten. Ausgehend von dem in Abbildung 2.4 dargestellten Lautsprecher-Raum-Mikrofon System lässt sich das gemessene Signal als

$$x(n) = s(n) * c(n) + \eta(n) \tag{2.3}$$

beschreiben, wobei das Rauschen $\eta(n)$ im Folgenden ignoriert wird.

Zur Messung der Raumimpulsantwort $c(n)$ kann das LRM System mit einem Einheitsimpuls angeregt werden; die Messung $x(n)$ entspricht dann der Raumimpulsantwort.

[1] http://home.tiscali.nl/ehabets/rir_generator.html, abgerufen am 26.09.2014

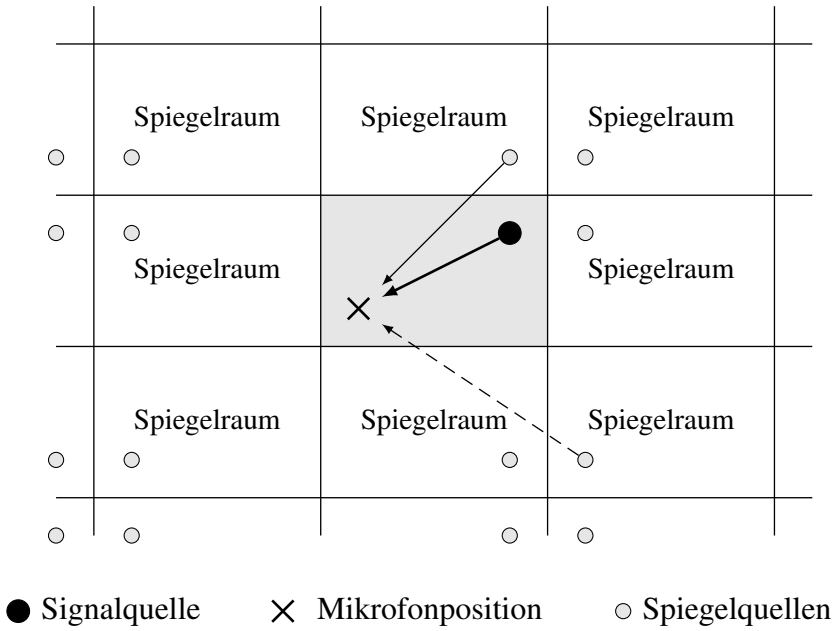

Abbildung 2.6: Spiegelquellenmodell nach Allen und Berkley [2]. Die Reflexionen der Schallwellen an den Wänden des Raumes werden durch mehrfaches *spiegeln* des Raumes in jede Raumdimension simuliert. Die Anzahl der Spiegelungen bestimmt die Anzahl der Reflexionen, die Laufzeit wird durch die Schallgeschwindigkeit und den euklidischer Abstand zwischen der Spiegelquelle und dem Empfänger bestimmt.

Tatsächlich wurden die ersten Messungen von Raumimpulsantworten durch die Erzeugung von impulsartigen Audiosignalen durchgeführt. Hierzu wurden beispielsweise Luftballons zum Platzen gebracht oder es wurde ein Pistolenschuss abgegeben [33]. Diese Methode ist ausreichend, um allgemeine raumakustische Parameter wie die Nachhallzeit zu bestimmen [98, 146, 147]. Die Anregung des LRM Systems mit einem impulsartigen Signal erweist sich in der Praxis als problematisch. So ist es zum einen nicht möglich, einen *perfekten* Impuls wiederzugeben, zum anderen ist diese Art der Messung hochgradig rauschanfällig [33]. Für die Messung von Raumimpulsantworten werden üblicherweise spezielle Anregungssignale mit bestimmten Eigenschaften verwendet. Die wichtigsten Anregungssignale werden im Folgenden kurz vorgestellt [9, 33, 157].

Folgen maximaler Länge Eine Folge maximaler Länge (englisch *maximum length sequence*, ML Sequenz) ist ein pseudozufälliges Binärsignal $b(n)$ der Länge $L_b = 2^N - 1$ mit $b(n) \in \{-1, 1\}$. Die Sequenz $b(n)$ wird hierbei durch ein N-Bit Schieberegister

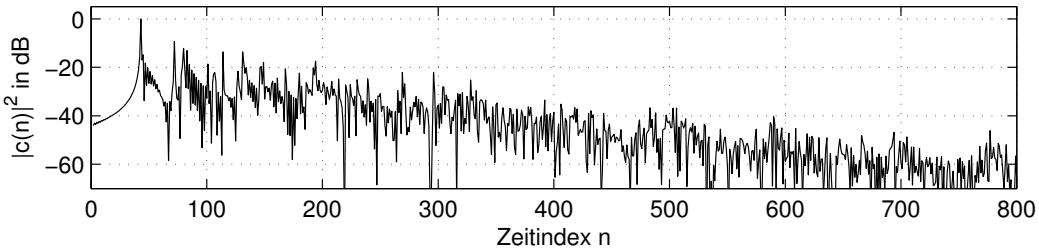

Abbildung 2.7: Nach der Spiegelquellenmethode [2] simulierte Raumimpulsantwort auf logarithmischer Skala.

generiert, welches mit einer XOR Operation verknüpft ist [33]. Sei die Sequenz $s(n)$ die periodisch wiederholte ML Sequenz. Wird $s(n)$ mit der ML Sequenz $b(n)$ korreliert, so gilt für die Kreuzkorrelationsfolge

$$r_{sb}(n) = \begin{cases} L_b, & \text{für } n = k\,L_b,\ k \in \mathbb{Z}, \\ -1, & \text{sonst.} \end{cases} \tag{2.4}$$

Wird L_b *ausreichend groß* gewählt, so approximiert die Kreuzkorrelationsfolge $r_{sb}(n)$ einen periodisch wiederholten und skalierten Einheitsimpuls [16, 162]:

$$r_{sb}(n) \approx \sum_k L_b\,\delta(n - k\,L_b)\,. \tag{2.5}$$

In Abbildung 2.8 (oben) ist ein Ausschnitt einer ML Sequenz der Länge $L_b = 255$ dargestellt; Abbildung 2.8 (unten) zeigt die Kreuzkorrelation der vierfach wiederholten ML Sequenz mit einer einzigen Periode. Ungeachtet der Randeffekte ist die impulsartige Struktur nach (2.5) deutlich zu erkennen.

Diese Besonderheit der ML Sequenzen wird für die Messung der Raumimpulsantwort ausgenutzt. Das zur Messung verwendete Anregungssignal $s(n)$ besteht dann aus einer periodischen Wiederholung der ML Sequenz $b(n)$. Wird das aufgenommene Mikrofonsignal $x(n)$ mit einer Periode der ML Sequenz korreliert, so ergibt sich

$$\begin{aligned} r_{xb}(n) &= x(n) * b(-n) = (s(n) * c(n)) * b(-n) \\ &\approx \sum_k L_b\,\delta(n - k\,L_b) * c(n)\,. \end{aligned} \tag{2.6}$$

Die Kreuzkorrelationsfolge $r_{xb}(n)$ des gemessenen Signals $x(n)$ mit einer Periode der ML Sequenz $b(n)$ enthält also periodisch wiederholt die um L_b skalierte Raumim-

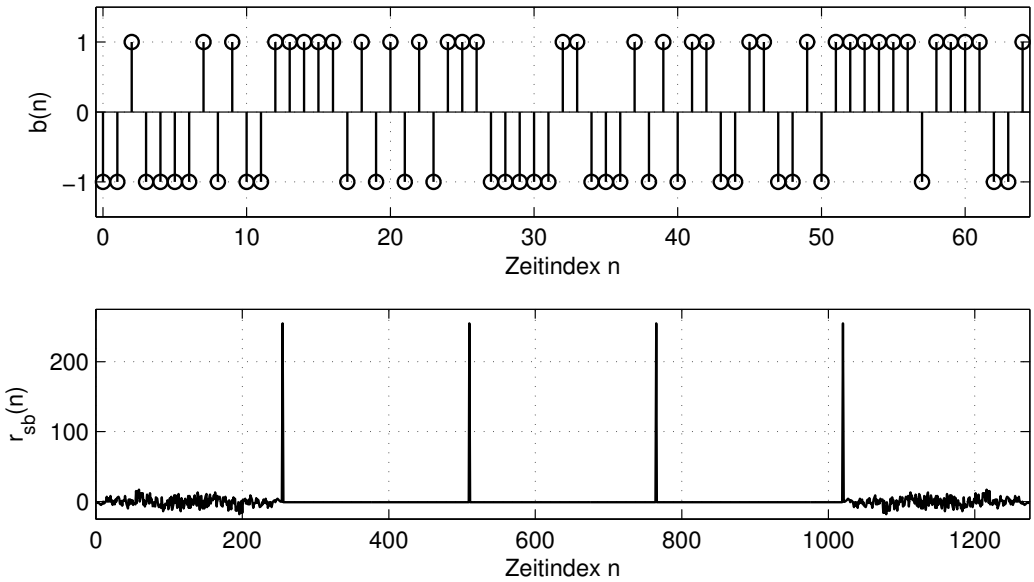

Abbildung 2.8: (Oben) Ausschnitt einer ML Sequenz $b(n)$ der Länge $L_b = 255$. (Unten) Kreuzkorrelation der vierfach wiederholten ML Sequenz mit einer einzigen Periode.

pulsantwort $c(n)$. In Abbildung 2.9 (oben) ist exemplarisch das Mikrofonsignal $x(n)$ bei Anregung mit einer periodisch wiederholten ML Sequenz dargestellt; Abbildung 2.9 (unten) zeigt die Kreuzkorrelationsfolge des Mikrofonsignals mit einer Periode der ML Sequenz dar. Das in Abbildung 2.9 (unten) dargestellte Signal enthält, wie in Gleichung (2.6) notiert, die periodisch wiederholte Raumimpulsantwort $c(n)$.

Eine Schätzung der Raumimpulsantwort $c(n)$ wird letztendlich durch eine Mittelung der einzelnen Perioden aus der Sequenz $r_{xb}(n)$ gewonnen [16, 33, 162].

Expenential Sine-Sweeps Zur Messung von Raumimpulsantworten werden heutzutage nahezu ausschließlich sogenannte *exponential sine sweeps* (ESS) verwendet [32, 33]. Ein ESS ist ein sinusförmiges Signal mit einer über die Zeit exponentiell ansteigenden Frequenz. In Abbildung 2.10 ist ein ESS im Zeitbereich (oben) sowie im Zeitfrequenzbereich (mitte) dargestellt; die exponentiell ansteigende Frequenz ist deutlich zu erkennen. Eine besondere Eigenschaft des ESS ist es, dass ein sogenanntes *inverses* Signal angegeben werden kann [32]. Bezeichnet $s(n)$ einen ESS und $s^{\text{inv.}}(n)$ das zu $s(n)$ inverse Signal, so gilt

$$s(n) * s^{\text{inv.}}(n) \approx \alpha\, \delta(n - n_0)\,, \tag{2.7}$$

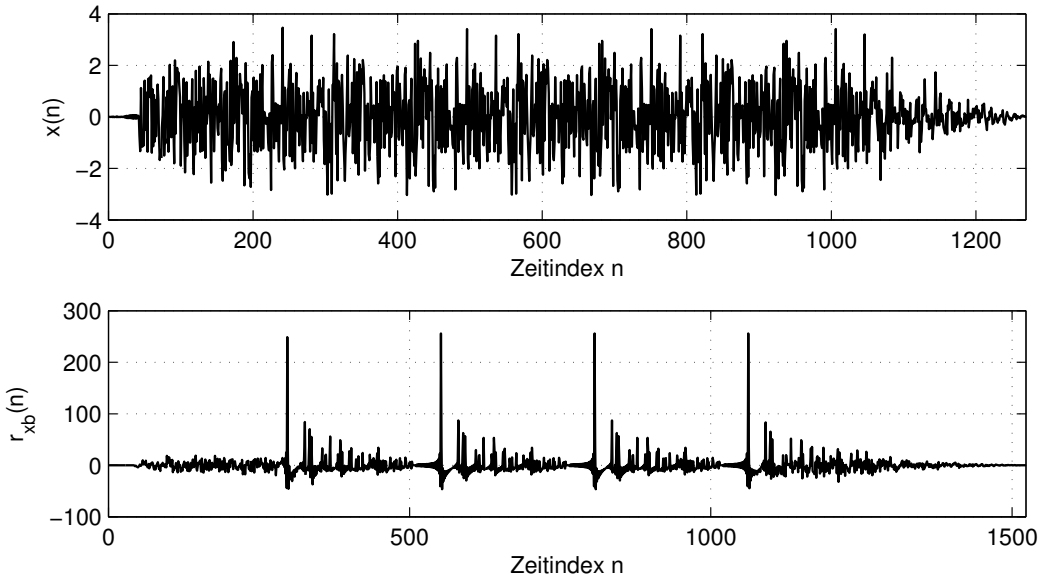

Abbildung 2.9: (Oben) Mikrofonsignal bei Anregung mit einer periodisch wiederholten ML Sequenz. (Unten) Kreuzkorrelationsfolge des Mikrofonsignals mit einer Periode der ML Sequenz.

wobei n_0 eine beliebige Verzögerung und α ein beliebiger Skalierungsfaktor ungleich null ist. In Abbildung 2.10 (unten) ist das Faltungsprodukt des ESS mit dem inversen ESS dargestellt.

Wird ein LRM System mit einem ESS angeregt, so kann durch Faltung des gemessenen Mikrofonsignals $x(n)$ mit dem zu dem ESS inversen Signal die Raumimpulsantwort ermittelt werden. Es gilt

$$x(n) * s^{\text{inv.}}(n) = s(n) * c(n) * s^{\text{inv.}}(n) \approx \alpha\delta(n - n_0) * c(n)\,. \tag{2.8}$$

Die Berechnungsvorschrift für einen ESS ist nach [32] durch

$$s(n) = \sin\left(\frac{\omega_L(L_s - 1)}{\log(\omega_U/\omega_L)} \cdot \left(e^{\frac{n}{L_s-1}\log(\omega_U/\omega_L)} - 1\right)\right) \tag{2.9}$$

gegeben, wobei L_s die Länge des Signals ist und ω_L und ω_U die untere beziehungsweise die obere Grenzfrequenz bezeichnen. Das zu (2.9) inverse Signal ist nach [32] durch

$$s^{\text{inv.}}(n) = s(L_s - 1 - n) \cdot \left(\frac{\omega_U}{\omega_L}\right)^{\frac{-n}{L_s-1}} \tag{2.10}$$

gegeben.

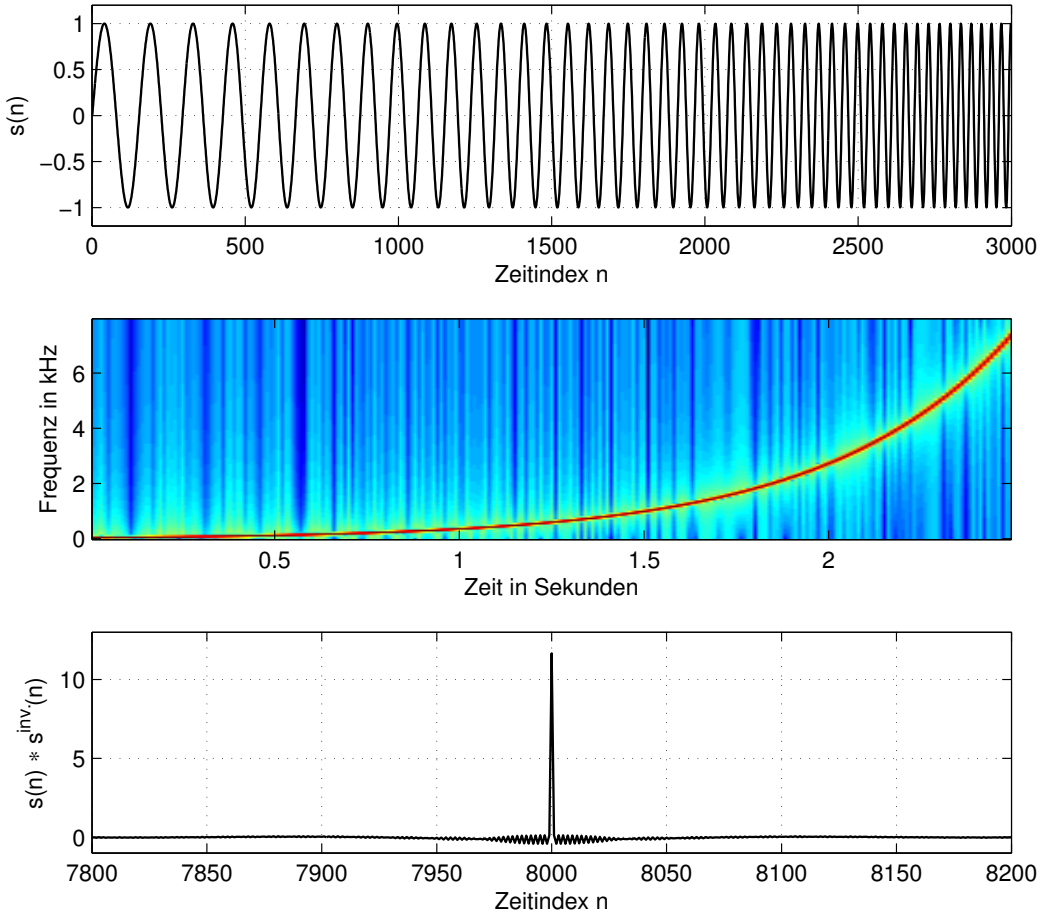

Abbildung 2.10: Ein *exponential sine sweep* (ESS) im Zeitbereich (oben) und in der Zeitfrequenzdarstellung (mitte). Faltungsprodukt des ESS mit dem *inversen* ESS (unten).

Weitere Methoden zur Messung von Raumimpulsantworten Ist das Anregungssignal $s(n)$ des LRM Systems bekannt, jedoch nicht frei wählbar, so kann die Raumimpulsantwort nach dem Prinzip der akustischen Echokompensation (englisch *acoustic echo cancellation*, AEC) geschätzt werden [6, 7]. Systeme zur akustischen Echokompensation werden beispielsweise in Freisprecheinrichtungen zur Unterdrückung des Sprachsignals des entfernten Sprechers im Mikrofonsignal eingesetzt. Der generelle Aufbau eines AECs ist in Abbildung 2.11 dargestellt.

Durch Minimierung der Energie des Fehlersignals $e(n)$ wird die Raumimpulsantwort $c(n)$ adaptiv von dem AEC geschätzt [6].

Ist das Anregungssignal $s(n)$ unbekannt, so ist es mit $N > 1$ Mikrofonen dennoch möglich, die Raumimpulsantworten von dem Lautsprecher zu den einzelnen Mikrofo-

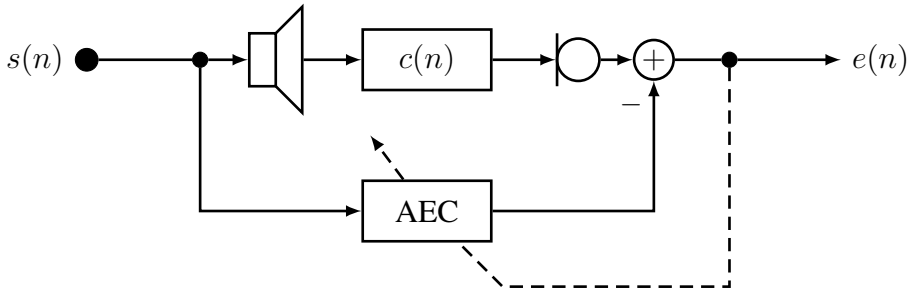

Abbildung 2.11: Aufbau eines Systems zur akustischen Echokompensation. Dir Raumimpulsantwort wird adaptiv von dem AEC Filter approximiert.

nen zu schätzen. Der generelle Aufbau eines solchen Systems ist in Abbildung 2.12 dargestellt.

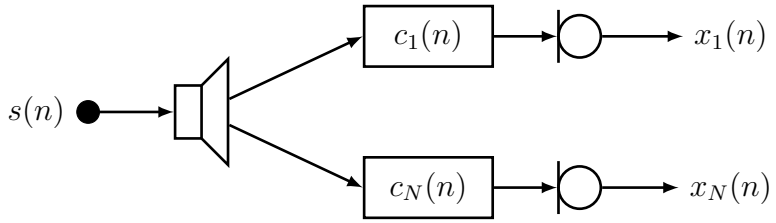

Abbildung 2.12: Aufbau eines Systems zur mehrkanaligen blinden Systemschätzung. Die Raumimpulsantworten $c_\ell(n)$, $1 \leq \ell \leq N$, können ohne Kenntnis des Anregungssignals aus den Mikrofonsignalen geschätzt werden.

Die einzelnen Raumimpulsantworten $c_\ell(n)$, $1 \leq \ell \leq N$, können ohne Kenntnis des Anregungssignals $s(n)$ lediglich aus den Mikrofonsignalen $x_1(n)$ bis $x_N(n)$ geschätzt werden [54]. Algorithmen zur mehrkanaligen blinden Systemidentifikation basieren üblicherweise auf dem Zusammenhang

$$x_j(n) * c_i(n) = x_i(n) * c_j(n)\,, \quad 1 \leq i, j \leq N, \;\; j \neq i. \tag{2.11}$$

Auf Grundlage von (2.11) wurden unterschiedliche Zielfunktionen mit unterschiedlichen Nebenbedingungen im Zeit- und/oder im Frequenzbereich formuliert [54, 56, 58, 65–67].

2.2 Temporale Maskierungseigenschaften des menschlichen Hörapparates

Nimmt ein menschlicher Zuhörer ein Schallereignis wahr, so *maskiert*, also verdeckt, dieses kurz darauffolgende Schallereignisse mit geringerem Pegel. Dies bedeutet, dass

nachfolgende Schallereignisse von einem menschlichen Zuhörer nur wahrgenommen werden, wenn ihr Pegel über einer bestimmten Schwelle, der sogenannten *temporalen Maskierungsschwelle*, liegt [24, 34, 35]. Der genaue Verlauf dieser Maskierungsschwelle ist sowohl signal- als auch pegelabhängig [24]. In [35] wurde eine durchschnittliche Maskierungsschwelle formuliert, die approximativ für unterschiedliche Anregungssignale gültig ist [68, 137, 172].

In Abbildung 2.13 ist der Verlauf der von Louis D. Fielder in [35] beschriebenen durchschnittlichen temporalen Maskierungsschwelle dargestellt; das Schallereignis (rot) findet zum Zeitpunkt $t = 0\,\mathrm{s}$ statt. Die temporale Maskierung beginnt $4\,\mathrm{ms}$ nach dem Schallereignis bei $-10\,\mathrm{dB}$ und fällt mit etwa $35\,\mathrm{dB}$ pro Dekade ab.

Der temporale Maskierungseffekt wird in aktuellen Verfahren zur kanalbasierten Raumhallkompensation ausgenutzt [120].

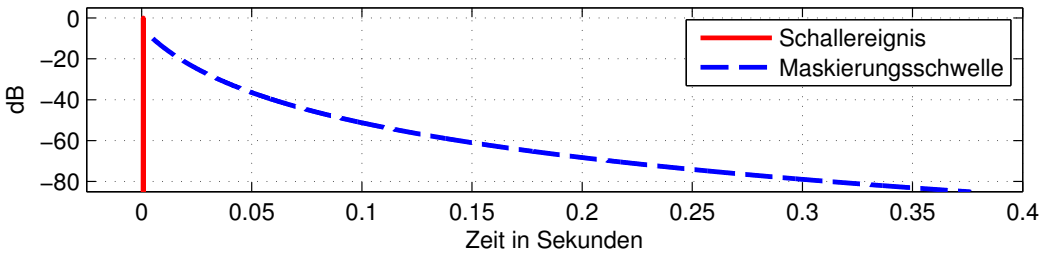

Abbildung 2.13: Durschnittliche temporale Maskierungsschwelle des menschlichen auditorischen Systems nach Louis D. Fielder [34, 35]. Die Maskierungsschwelle (blau) wird durch das Schallereignis (rot) induziert. Nachfolgende Schallereignisse sind nur wahrnehmbar, sofern ihr Pegel über der Maskierungsschwelle liegt.

Frequenzmaskierung des menschlichen Hörsystems Von großer Bedeutung für die verlustbehaftete Kompression von Audiosignalen ist der frequenzabhängige Maskierungseffekt [50, 172]. Ein Ton mit einer bestimmten Mittenfrequenz induziert, analog zur temporalen Maskierungsschwelle im Zeitbereich, eine Maskierungsschwelle im Frequenzbereich. Gleichzeitig auftretende Schallereignisse einer anderen Frequenz werden nur wahrgenommen, wenn ihre Energie über der frequenzabhängigen Maskierungsschwelle liegt. Die frequenzabhängigen Maskierungseffekte werden in den im Rahmen dieser Arbeit entwickelten Verfahren nicht ausgenutzt und sind an dieser Stelle lediglich der Vollständigkeit halber angesprochen. Eine exemplarische Darstellung des frequenzabhängigen Maskierungseffektes ist in Abbildung 2.14 gegeben.

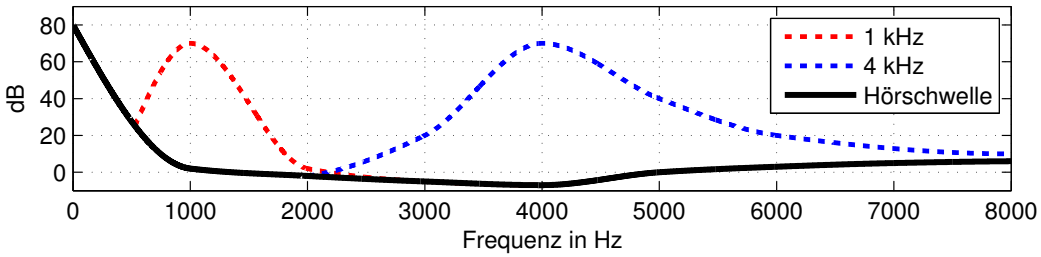

Abbildung 2.14: Darstellung der Hörschwelle (schwarz). Exemplarische Darstellung der Maskierungsschwellen für einen Ton der Frequenz $1\,\mathrm{kHz}$ (rot) und einen Ton der Frequenz $4\,\mathrm{kHz}$ (blau).

2.3 Bewertung akustischer Übertragungskanäle

Ein akustischer Kanal wird vollständig durch seine Impulsantwort beschrieben. Die Bewertung des Kanals auf Grundlage aller zur Verfügung stehenden Parameter (zum Beispiel alle Koeffizienten der Impulsantwort) ist schwer zu greifen. Deswegen ist es nötig, die große Anzahl an Parametern zu einigen wenigen, quantitativen Maßen zu kondensieren [98].

Zur Bewertung akustischer Übertragungssysteme stehen im Allgemeinen zwei unterschiedliche Methoden zur Verfügung. Zum einen kann der akustische Kanal auf Grundlage der Impulsantwort beziehungsweise Übertragungsfunktion bewertet werden [98], zum anderen kann beispielsweise ein durch den Kanal übertragenes Sprachsignal mit dem Originalsignal verglichen werden. Diese beiden Ansätze zur Bewertung von akustischen Übertragungssystemen werden als *kanal-* beziehungsweise *signalbasierte* Maße bezeichnet [44, 46, 102]. Signalbasierte Kriterien werden oftmals zur Bewertung von Kompressionsalgorithmen verwendet, die auf einer nichtlinearen Manipulation des Signals basieren.

In [44] und [46] wurde gezeigt, dass kanalbasierte Maße generell eine höhere Korrelation zur wahrgenommenen Qualität aufweisen als die signalbasierten Maße. Im Folgenden werden einige der von Goetze et al. in [44] untersuchten Maße vorgestellt.

Im Verlauf dieser Arbeit wurden zwei weitere Maße erarbeitet, die in Abschnitt 6.1 vorgestellt werden.

2.3.1 Maße zur Bewertung der Zeitbereichsdarstellung

Die Maße zur Bewertung einer Raumimpulsantwort im Zeitbereich quantisieren oftmals die Verteilung der Energie über den zeitlichen Verlauf der Impulsantwort [98, 133].

Hierbei wird die Energie der Direktschallkomponente und/oder der frühen Reflexionen mit der Gesamtenergie der Raumimpulsantwort oder der Energie der späten Reflexionen verglichen. Die wichtigsten Maße sind im Folgenden vorgestellt.

Nachhallzeit Die Nachhallzeit eines Raumes ist die Zeit, zu der die Energie der Zeitkoeffizienten der Raumimpulsantwort im Verhältnis zur Energie der Direktschallkomponente um $-60\,\mathrm{dB}$ abgeklungen ist [98]. Die Nachhallzeit wird mit T_{60} notiert. Aus einer gemessenen Raumimpulsantwort kann die Nachhallzeit beispielsweise durch die Rückintegrationsmethode nach Schroeder bestimmt werden [146].

In Abbildung 2.15 ist eine mit der Spiegelquellenmethode nach [2] simulierte Raumimpulsantwort dargestellt; für die Simulation wurde eine Nachhallzeit von $T_{60} = 0.3$ Sekunden gewählt.

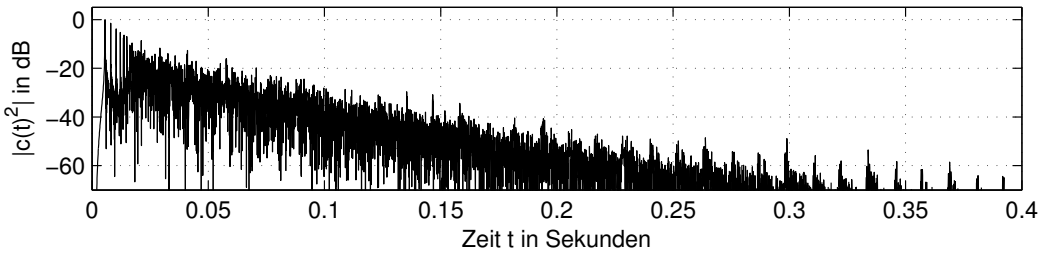

Abbildung 2.15: Simulierte Raumimpulsantwort mit einer Nachhallzeit von $T_{60} = 0.3$ Sekunden.

Deutlichkeit Ein erster Versuch, die Sprachverständlichkeit (genauer Silbenverständlichkeit, [144]) auf Grundlage einer Impulsantwort zu quantifizieren, wurde von Meyer und Thiele in [121] unternommen [98]. Das Maß wurde als „Deutlichkeit" (englisch *definition*) bezeichnet und ist ursprünglich als

$$\mathrm{D} = \frac{\int_0^{50\,\mathrm{ms}} c^2(t)\ dt}{\int_0^{\infty} c^2(t)\ dt} \cdot 100\ \% \tag{2.12}$$

definiert worden, wobei $c(t)$ die Impulsantwort des akustischen Kanals bezeichnet. Unter der Annahme, dass der Direktschall zur Zeit $t = 0\,\mathrm{s}$ eintrifft, misst die Deutlichkeit nach [121] das Verhältnis der Energie in den ersten $50\,\mathrm{ms}$ (beginnend mit dem Direktschall) der Raumimpulsantwort zur Gesamtenergie der Raumimpulsantwort [133]. Ist die Ge-

samtenergie der Impulsantwort $c(t)$ in den ersten $50\,\mathrm{ms}$ konzentriert, so nimmt (2.12) den Maximalwert von $100\,\%$ an.

Für eine diskrete Impulsantwort $c(n)$ der Länge L_c, die mit der Frequenz f_s abgetastet wurde berechnet sich die Deutlichkeit als

$$\mathrm{D} = \frac{\sum\limits_{n=n_0}^{n_0+f_s 50\,\mathrm{ms}} c^2(n)}{\sum\limits_{n=0}^{L_c-1} c^2(n)}, \tag{2.13}$$

wobei n_0 der Zeitindex der Direktschallkomponente ist. Im Gegensatz zu (2.12) ist eine Skalierung in Prozent heute eher unüblich. Als Abwandlungen von (2.13) finden sich in der Literatur noch andere Summationsgrenzen. So ist es durchaus üblich, nicht die Energie in den ersten $50\,\mathrm{ms}$, sondern in den ersten $30\,\mathrm{ms}$ oder $80\,\mathrm{ms}$ ins Verhältnis zur Gesamtenergie zu setzen [44, 46]. Um Unsicherheiten zu vermeiden wird der Zeitbereich oftmals als Subskript angegeben; somit würde (2.13) auch als D_{50} bezeichnet werden, da die Energie über $50\,\mathrm{ms}$ integriert wird.

Klarheit Die Klarheit (auch „Klarheitsmaß", englisch *clarity*) ist formal ähnlich wie die Deutlichkeit (2.12) definiert: Auch das Klarheitsmaß bewertet die temporale Verteilung der Energie einer Impulsantwort [98]. Die Klarheit wurde ursprünglich von Reichardt et al. in [141] eingeführt, um die Transparenz von Musik in einem Konzertsaal zu quantifizieren. Die Klarheit nach [141] ist für eine Raumimpulsantwort $c(t)$ als

$$\mathrm{C} = 10\log_{10}\left(\frac{\int\limits_{0}^{80\,\mathrm{ms}} c^2(t)\;dt}{\int\limits_{80\,\mathrm{ms}}^{\infty} c^2(t)\;dt}\right)\;[\mathrm{dB}] \tag{2.14}$$

definiert. Unter der Annahme, dass der Direktschall bei $t = 0\,\mathrm{s}$ eintrifft, quantifiziert (2.14) also die Energie in den ersten $80\,\mathrm{ms}$ der Impulsantwort im Verhältnis zur Energie im verbleibenden Teil der Impulsantwort.

Die Integrationsgrenze von $80\,\mathrm{ms}$ im Gegensatz zu den $50\,\mathrm{ms}$ aus (2.12) wird dadurch motiviert, dass ein menschlicher Zuhörer den Nachhall bei Musikdarbietungen üblicherweise weniger gut wahrnehmen kann (beziehungsweise als weniger *störend* empfindet) als bei Sprachsignalen [98]. Die Klarheit nach (2.14) für Konzertsäle liegt üblicherweise zwischen $-3\,\mathrm{dB}$ und $5\,\mathrm{dB}$ [98].

Für eine diskrete Impulsantwort $c(n)$ der Länge L_c wird die Klarheit als

$$\mathrm{C} = 10\log_{10}\left(\frac{\sum\limits_{n=n_0}^{n_0+f_s 80\,\mathrm{ms}} c^2(n)}{\sum\limits_{n=0}^{n_0-1} c^2(n) + \sum\limits_{n=n_0+f_s 80\,\mathrm{ms}+1}^{L_c-1} c^2(n)}\right) \quad [\mathrm{dB}] \tag{2.15}$$

berechnet, wobei n_0 der diskrete Zeitindex der Direktschallkomponente ist und f_s die Abtastrate bezeichnet.

Als Abwandlungen von (2.15) finden sich in der Literatur noch andere Summationsgrenzen. So ist es durchaus üblich, nicht die Energie in den ersten $80\,\mathrm{ms}$, sondern in den ersten $30\,\mathrm{ms}$ oder $50\,\mathrm{ms}$ ins Verhältnis zur Energie der restlichen Impulsantwort zu setzen [44, 46]. Um Unsicherheiten zu vermeiden wird der Zeitbereich oftmals als Subskript angegeben; somit würde (2.15) auch als C_{80} bezeichnet werden, da die Energie über $80\,\mathrm{ms}$ integriert wird.

Direktschall-zu-Nachhall Verhältnis Das Direktschall-zu-Nachhall Verhältnis (englisch *direct-to-reverberation ratio*, DRR) ist nach [161] als das Verhältnis der Energie der Direktschallkomponente zur Energie der restlichen Impulsantwort definiert. Da bei einer diskreten Impulsantwort die Direktschallkomponente nicht notwendigerweise auf exakt einen Abtastwert abgebildet wird, wird ein kleiner Bereich um den Höchstpunkt (englisch *main peak*) der Impulsantwort integriert [46, 52]; dieser Bereich wird als „Direktpfad" bezeichnet.

Nach [46] wird das Direktschall-zu-Nachhall Verhältnis einer diskreten, mit der Frequenz f_s abgetasteten Raumimpulsantwort $c(n)$ als

$$\mathrm{DRR} = 10\log_{10}\left(\frac{\sum\limits_{n=n_0-n_\Delta}^{n_0+n_\Delta} c^2(n)}{\sum\limits_{n=0}^{n_0-n_\Delta-1} c^2(n) + \sum\limits_{n=n_0+n_\Delta+1}^{L_c-1} c^2(n)}\right) \quad [\mathrm{dB}] \tag{2.16}$$

berechnet, wobei n_0 die Position der Direktschallkomponente und n_Δ den Bereich des Direktpfades bestimmt; in [46] wird $n_\Delta = f_s 4\,\mathrm{ms}$ verwendet.

Zentrums-Zeit Die Zentrums-Zeit (englisch *center time*, CT) ist der Schwerpunkt der quadrierten Impulsantwort [97, 98, 173] beziehungsweise das erste Moment der

quadrierten Impulsantwort. Die Zentrums-Zeit einer kontinuierlichen Impulsantwort $c(t)$ berechnet sich als

$$\mathrm{CT} = \frac{\int_0^\infty t c^2(t)\ dt}{\int_0^\infty c^2(t)\ dt}\ \ [\mathrm{s}]\,. \tag{2.17}$$

Entsprechend berechnet sich die Zentrums-Zeit für eine diskrete Impulsantwort $c(n)$ der Länge L_c als

$$\mathrm{CT} = \left(\frac{\sum_{n=0}^{L_c-1} n c^2(n)}{\sum_{n=0}^{L_c-1} c^2(n)} \right) f_s^{-1}\ \ [\mathrm{s}]\,, \tag{2.18}$$

wobei f_s die Abtastfrequenz bezeichnet [46].

Im Vergleich zu der Deutlichkeit D_{50} (2.13) oder der Klarheit C_{80} (2.15) zeichnet sich die Zentrums-Zeit dadurch aus, dass keine willkürliche Aufteilung der Impulsantwort in unterschiedliche Epochen vorgenommen wird [173].

Die Zentrums-Zeit ist stark negativ mit der Sprachverständlichkeit korreliert [97, 98]. Für die Darbietung von Sprachsignalen ist eine Zentrums-Zeit von $60\,\mathrm{ms} \leq \mathrm{CT} \leq 80\,\mathrm{ms}$ anzustreben [173].

2.3.2 Bewertung der Frequenzbereichsdarstellung

Um eine Übertragung ohne spektrale Verzerrungen zu garantieren, sollte der akustische Kanal einen *flachen* Betragsfrequenzgang aufweisen [44]. Zur Bewertung der spektralen Färbung des Signals durch den akustischen Kanal werden beispielsweise die Varianz des logarithmierten Betragsfrequenzgangs [127] oder der Quotient aus arithmetischem und geometrischem Mittel des quadrierten Betragsfrequenzgangs verwendet [105].

Maß zur Bewertung der spektralen Flachheit Das Verhältnis von geometrischem zu arithmetischem Mittel wurde bereits in [19] und [59] verwendet, um den Streubereich einer Datenreihe zu messen [105]. Sind alle Werte nahezu identisch, so ist der Streubereich *klein* und der Quotient strebt gegen eins; sind die Werte sehr un-

terschiedlich, so strebt der Quotient gegen null [105]. In Abbildung 2.16 sind zwei Datenreihen dargestellt. Mit

$$\mathrm{FM}_{x(n)} = \frac{\left(\prod_{n=0}^{L_x-1} |x(n)|\right)^{\frac{1}{L_x}}}{\frac{1}{L_x}\sum_{n=0}^{L_x-1} |x(n)|} \tag{2.19}$$

als Quotient des geometrischen und des arithmetischen Mittels der Beträge der Datenreihe $x(n)$ der Länge L_x gilt für die in Abbildung 2.16 dargestellten Datenreihen $\mathrm{FM}_{x_1(n)} = 0.48$ und $\mathrm{FM}_{x_2(n)} = 0.69$.

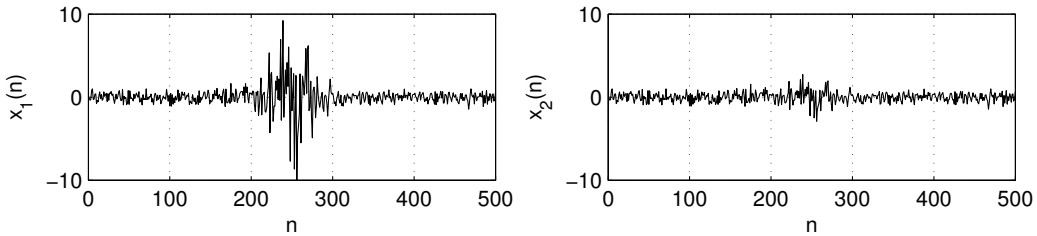

Abbildung 2.16: Zwei zufällige Datenreihen. Der Quotient aus geometrischem und arithmetischem Mittel der Beträge beträgt für $x_1(n)$ einen Wert von 0.48, für $x_2(n)$ hat der Quotient einen Wert von 0.69.

Das Maß zur Bewertung der spektralen Flachheit (englisch *spectral flatness measure*, SFM) ist als Quotient des geometrischen Mittels und des arithmetischen Mittels der spektralen Leistungsdichte eines Signals definiert [51, 70, 105]. Die Definition des SFM für ein diskretes Signal $s(n)$ nach [105] lautet

$$\mathrm{SFM}_{s(n)} = \frac{\exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(\widehat{P}(\omega)\right)\, d\omega\right]}{\frac{1}{2\pi}\int_{-\pi}^{\pi} \widehat{P}(\omega)\, d\omega}, \tag{2.20}$$

wobei $\widehat{P}(\omega) = \left|S(e^{j\omega})\right|^2$ das Betragsquadrat der zeitdiskreten Fourier-Transformierten des Signals $s(n)$ ist.

Spektrale Flachheit eines Systems Im Folgenden wird das SFM eines Systems als Quotient des geometrischen Mittels des quadrierten Betragsfrequenzgangs durch

das arithmetische Mittel des quadrierten Betragsfrequenzgangs verstanden. Für eine Impulsantwort $g(n)$ wird das *spectral flatness measure* als

$$\mathrm{SFM}_{g(n)} = \frac{\left(\prod\limits_{k=1}^{K} \left|G(e^{j\omega_k})\right|^2\right)^{\frac{1}{K}}}{\frac{1}{K}\sum\limits_{k=1}^{K} \left|G(e^{j\omega_k})\right|^2} \tag{2.21}$$

definiert, wobei ω_k, $k = 1, \ldots, K$, die betrachteten diskreten Frequenzen sind. Die Übertragungsfunktion $G(e^{j\omega_k})$ wird hierbei als diskrete Fourier-Transformation der Länge K der Impulsantwort $g(n)$ berechnet.

2.4 Zusammenfassung

In diesem Kapitel wurden die Grundlagen der Schallausbreitung in geschlossenen Räumen erklärt. Des Weiteren wurde die Beschreibung akustischer Übertragungskanäle innerhalb eines Raumes durch die Raumimpulsantwort erläutert und es wurde die Simulation und Messung von Raumimpulsantworten erklärt. Abschließend wurde ein kurzer Überblick über die Maskierungseigenschaften des menschliche Hörsystems gegeben und es wurden mehrere Maße zur quantitativen Bewertung von akustischen Kanälen vorgestellt.

3 Kanalbasierte Raumhallkompensation

In diesem Kapitel werden unterschiedliche Methoden zur kanalbasierten Raumhallkompensation vorgestellt. Des Weiteren werden die Vor- und Nachteile der einzelnen Verfahren diskutiert.

Eine der größten Herausforderungen der kanalbasierten Raumhallkompensation ist der Entwurf von Filtern, die robust gegenüber Kanalfehlschätzungen sowie räumlicher Fehlpositionierung sind. Diese Problematik wird in Abschnitt 3.4 näher betrachtet. Die eng mit der Kanalkompensation verwandte Problemstellung der Übersprechkompensation wird in Abschnitt 3.5 eingeführt.

3.1 Übersicht

Generell existieren zwei Klassen an Verfahren zur Kanalkompensation. Zum einen muss ein zu approximierendes Zielsystem vorgegeben werden, zum anderen können lediglich gewisse Anforderungen an das Gesamtsystem gestellt werden. Die Güte der Entzerrung ist von der Wahl des zu approximierenden Zielsystems abhängig [96]. Für dieses gibt es eine Vielzahl an freien Parametern, wie beispielsweise eine Verzögerung oder bestimmte Anforderungen an den Frequenzgang des Zielsystems. Obwohl es erste Ansätze gibt, das gewünschte Zielsystem zumindest teilweise gemeinsam mit den Filterkoeffizienten zu optimieren [83], ist die automatische Wahl eines *guten* Zielsystems im Allgemeinen nicht möglich [96]. Der Fokus dieser Arbeit liegt auf Verfahren zur Formung von Raumimpulsantworten, bei denen das gewünschte Verhalten des Gesamtsystems durch die Wahl zweier Gewichtsfenster vorgegeben wird. Verfahren zur Formung werden in Abschnitt 3.3 besprochen. Die in Abschnitt 3.2 aufgeführten Verfahren zu Kanalkompensation mit vorgegebenem Zielsystem werden lediglich der Vollständigkeit halber vorgestellt.

Viele Verfahren zur Kanalformung beziehungsweise Kanalverkürzung wurden ursprünglich für die Datenübertragung entwickelt [31]. Im Gegensatz zu den üblicherweise *kurzen* Impulsantworten in der elektronischen Datenübertragung (meistens unter 200

Koeffizienten, zum Beispiel [15, 20]) sind akustische Kanäle üblicherweise sehr lang (über 2000 Koeffizienten, zum Beispiel [120]).

Abbildung 3.1 zeigt den generellen Aufbau der im Rahmen dieser Arbeit betrachteten Szenarien. Ein Kompensationsfilter $h(n)$ ist derart zu entwerfen, dass das Faltungsprodukt $g(n)$ aus der Raumimpulsantwort $c(n)$ und $h(n)$ *bessere* akustische Eigenschaften hat als die Raumimpulsantwort selber. Aufgrund der Kommutativität der Faltung kann das Kompensationsfilter sowohl zur Vorfilterung vor der Lautsprecherwiedergabe, als auch zur Nachbearbeitung von aufgenommenen Audiosignalen verwendet werden.

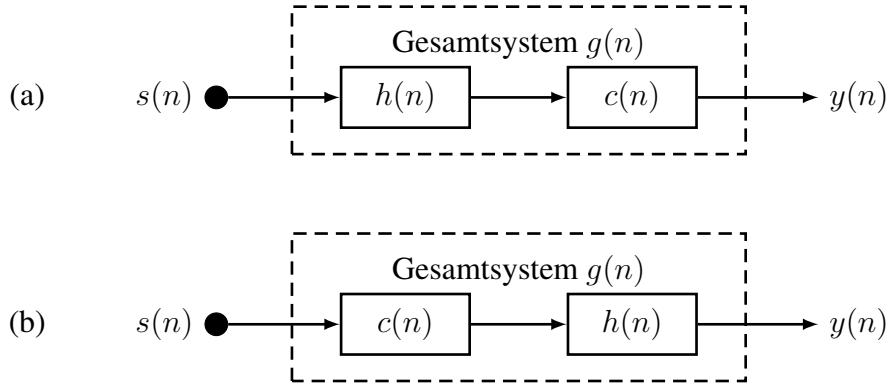

Abbildung 3.1: In den dargestellten Szenarien wird der Entzerrer als Filter vor dem Lautsprecher (a) beziehungsweise als Filter zur Nachbearbeitung des Mikrofonsignals (b) verwendet. Aufgrund der Kommutativität der Faltung [119] sind beide Szenarien mathematisch äquivalent.

Inversion von Übertragungssystemen Das Problem der Kanalkompensation ist eng verwandt mit dem Problem der Systeminversion. Die Invertierung führt zu einer vollständigen Kompensation des Systems. Das zu $C(z)$ inverse System ist durch

$$H(z) = \frac{1}{C(z)} \tag{3.1}$$

gegeben. Mit (3.1) gilt $H(z)\,C(z) = 1$, oder im Zeitbereich $h(n) * c(n) = \delta(n)$.

Diese mathematisch *einfache* Lösung ist für die Kompensation akustischer Kanäle nicht anwendbar. Raumimpulsantworten sind üblicherweise nicht minimalphasig [35, 127, 130], weswegen das inverse System nicht-kausal und nicht-stabil werden würde und damit praktisch nicht realisierbar ist [119].

In einer aktuellen Veröffentlichung wird der Ansatz der direkten Systeminvertierung nach (3.1) dennoch verfolgt [92]. Die Invertierung erfolgt hierbei im Kurzzeit-Fourier-

Raum. Die aus dem antikausalen Charakter des inversen Systems $H(z)$ resultierenden Artefakte (Prä-Echos) werden in einem zweiten Verarbeitungsschritt durch ein spektrales Subtraktionsverfahren [39, 86, 99] aus dem Signal entfernt [92]. Das Verfahren aus [92] ist somit der Klasse der nichtlinearen Verfahren zuzuordnen und kann lediglich zur Nachbearbeitung von Aufnahmen verwendet werden, nicht aber zur Vorfilterung.

In den folgenden Abschnitten wird ein Überblick über mehrere Verfahren zur Kanalkompensation mit linearen FIR Filtern gegeben.

3.2 Kanalkompensation mit gegebenem Zielsystem

In diesem Abschnitt wird der Filterentwurf mit einem vorgegebenen Zielsystem behandelt. Der Signalflussgraph für dieses Szenario ist in Abbildung 3.2 dargestellt. Für die vorgestellten Verfahren wird angenommen, dass die tatsächliche Raumimpulsantwort $c(n)$ bekannt ist und fehlerfrei geschätzt wurde.

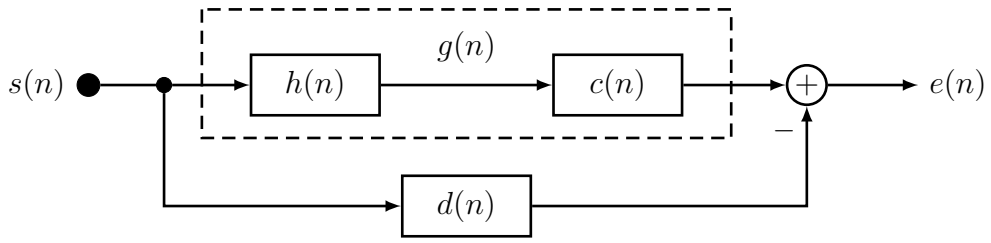

Abbildung 3.2: Signalflussgraph für die Kompensation eines bekannten Kanals $c(n)$ durch ein Filter $h(n)$. Das Faltungsprodukt $g(n) = c(n) * h(n)$ soll das vorgegebene Zielsystem $d(n)$ in optimaler Weise approximieren.

In Abbildung 3.2 bezeichnet $c(n)$ die zu kompensierende Raumimpulsantwort und $h(n)$ das zu entwerfende Filter. Die globale Impulsantwort $g(n) = h(n) * c(n)$ soll in *optimaler* Weise das vorgegebene Zielsystem $d(n)$ approximieren. Hierzu wird ein vom Fehlersignal $e(n)$ abhängiges Kriterium in geeigneter Weise minimiert. Das Fehlersignal $e(n)$ ist durch

$$e(n) = s(n) * \underbrace{(h(n) * c(n))}_{g(n)} - s(n) * d(n) = s(n) * (g(n) - d(n)) \tag{3.2}$$

gegeben. Das gewünschte Zielsystem $d(n)$ wird üblicherweise als verzögerter beziehungsweise als verzögerter und bandpass gefilterter Einheitsimpuls gewählt [85].

3.2.1 Methode des kleinsten quadratischen Fehlers

Das in diesem Abschnitt vorgestellte Verfahren wurde von Falconer und Magee [31] für die Kanalverkürzung in der Datenübertragung untersucht. Es basiert auf der Minimierung des quadratischen Fehlers zwischen dem Faltungsprodukt aus dem Entzerrer $h(n)$ mit der Kanalimpulsantwort $c(n)$ und einem vorgegebenen Zielsystem $d(n)$. Die Methode des kleinsten quadratischen Fehlers basiert somit auf der Minimierung der Energie des Fehlersignals $e(n)$, welches durch (3.2) gegeben ist [130]. Das Minimierungsproblem lautet

$$\underset{h(n)}{\text{minimize}} \quad : \quad \mathrm{E}\big\{e^2(n)\big\}\,. \tag{3.3}$$

Für ein Filter der Länge L_h sind also die Filterkoeffizienten $h(n)$, $0 \leq n \leq L_h - 1$, derart zu wählen, dass die Energie des Fehlersignals $e(n)$ minimal wird. Die Raumimpulsantwort $c(n)$ besitzt die Länge L_c und wird als bekannt angenommen; die globale Impulsantwort $g(n)$ sowie das zu approximierende Zielsystem $d(n)$ haben die Länge $L_g = L_c + L_h - 1$.

3.2.1.1 Herleitung

Zur Lösung des in (3.3) dargestellten Optimierungsproblems werden die beteiligten Komponenten als Vektoren und Matrizen geschrieben. Der Vektor $\boldsymbol{h}$ enthält die zu optimierenden Filterkoeffizienten und der Vektor $\boldsymbol{d}$ die des Zielsystems:

$$\boldsymbol{h} = \begin{bmatrix} h(0) \\ \vdots \\ h(L_h - 1) \end{bmatrix}, \qquad \boldsymbol{d} = \begin{bmatrix} d(0) \\ \vdots \\ d(L_g - 1) \end{bmatrix}. \tag{3.4}$$

Der Vektor $\boldsymbol{s}(n)$ enthält einen L_g Zeitkoeffizienten langen Abschnitt des Anregungssignals $s(n)$:

$$\boldsymbol{s}(n) = \begin{bmatrix} s(n) \\ s(n-1) \\ \vdots \\ s(n - L_g + 1) \end{bmatrix}. \tag{3.5}$$

Die Matrix $\boldsymbol{C}$ der Form

$$\boldsymbol{C} = \begin{bmatrix} c(0) & 0 & \cdots & 0 \\ c(1) & c(0) & 0 & \vdots \\ \vdots & x(1) & \ddots & 0 \\ c(L_c-1) & \vdots & \ddots & c(0) \\ 0 & c(L_c-1) & & c(1) \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & c(L_c-1) \end{bmatrix} \tag{3.6}$$

wird als Faltungsmatrix der Raumimpulsantwort $c(n)$ bezeichnet.

Mit (3.4), (3.5) und $\boldsymbol{C}$ als Faltungsmatrix (3.6) der Größe $L_g \times L_h$ der Raumimpulsantwort $c(n)$ kann das Fehlersignal (3.2) als

$$e(n) = \underbrace{\boldsymbol{s}^T(n)\,\boldsymbol{C}\boldsymbol{h}}_{s(n)*c(n)*h(n)} - \underbrace{\boldsymbol{s}^T(n)\,\boldsymbol{d}}_{s(n)*d(n)} \tag{3.7}$$

geschrieben werden. Entsprechend gilt für das in (3.3) formulierte Optimierungsproblem folgender äquivalenter Ausdruck in Matrix-Vektor-Notation:

$$\underset{\boldsymbol{h}}{\text{minimize}} \;:\; f(\boldsymbol{h})\,, \quad f(\boldsymbol{h}) = \mathrm{E}\Big\{\big(\boldsymbol{s}^T(n)\,\boldsymbol{C}\boldsymbol{h} - \boldsymbol{s}^T(n)\,\boldsymbol{d}\big)^2\Big\}\,. \tag{3.8}$$

Ausgeschrieben ergibt sich für die Energie des Fehlersignals

$$\mathrm{E}\big\{e^2(n)\big\} = \mathrm{E}\big\{\boldsymbol{h}^T\boldsymbol{C}^T\boldsymbol{s}(n)\,\boldsymbol{s}^T(n)\,\boldsymbol{C}\boldsymbol{h} - 2\boldsymbol{h}^T\boldsymbol{C}^T\boldsymbol{s}(n)\,\boldsymbol{s}^T(n)\,\boldsymbol{d} + \boldsymbol{d}^T\boldsymbol{s}(n)\,\boldsymbol{s}^T(n)\,\boldsymbol{d}\big\}\,. \tag{3.9}$$

Mit

$$\boldsymbol{R}_{\boldsymbol{ss}}(n) = \mathrm{E}\big\{\boldsymbol{s}(n)\,\boldsymbol{s}^T(n)\big\} \tag{3.10}$$

und der Aufteilung von (3.9) in stochastische (das Anregungssignal $s(n)$ betreffende) und deterministische (die Raumimpulsantwort $c(n)$ sowie den zu entwerfenden Entzerrer $h(n)$ betreffende) Komponenten ergibt sich

$$\mathrm{E}\big\{e^2(n)\big\} = \boldsymbol{h}^T\boldsymbol{C}^T\boldsymbol{R}_{\boldsymbol{ss}}(n)\,\boldsymbol{C}\boldsymbol{h} - 2\boldsymbol{h}^T\boldsymbol{C}^T\boldsymbol{R}_{\boldsymbol{ss}}(n)\,\boldsymbol{d} + \boldsymbol{d}^T\boldsymbol{R}_{\boldsymbol{ss}}(n)\,\boldsymbol{d}. \tag{3.11}$$

Mit der Annahme, dass $s(n)$ ein weißes Rauschsignal mit Varianz $\sigma^2_{s(n)} = 1$ ist gilt $\boldsymbol{R_{ss}}(n) = \boldsymbol{I}$, und die Kostenfunktion aus (3.8) vereinfacht sich zu

$$f(\boldsymbol{h}) = \boldsymbol{h}^T\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{h} - 2\boldsymbol{h}^T\boldsymbol{C}^T\boldsymbol{d} + \boldsymbol{d}^T\boldsymbol{d}. \tag{3.12}$$

Mit (3.12) kann das Optimierungsproblem aus (3.3) somit als

$$\underset{\boldsymbol{h}}{\text{minimize}} \;:\; f(\boldsymbol{h})\,, \quad f(\boldsymbol{h}) = \boldsymbol{h}^T\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{h} - 2\boldsymbol{h}^T\boldsymbol{C}^T\boldsymbol{d} + \boldsymbol{d}^T\boldsymbol{d} \tag{3.13}$$

geschrieben werden. Die Kostenfunktion $f(\boldsymbol{h})$ aus (3.13) ist quadratisch und besitzt genau ein globales Minimum. Zur Lösung von (3.13) werden die partiellen Ableitungen der Kostenfunktion bezüglich der Filterkoeffizienten $h(n)$, $0 \leq n < L_h$, gebildet und gleich null gesetzt.

In Vektor-Notation geschrieben ist der Gradient bezüglich des Zielvektors $\boldsymbol{h}$ gleich

$$\nabla_{\boldsymbol{h}} f(\boldsymbol{h}) = 2\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{h} - 2\boldsymbol{C}^T\boldsymbol{d}. \tag{3.14}$$

An der Stelle des globalen Minimums von $f(\boldsymbol{h})$ muss der Gradient $\nabla_{\boldsymbol{h}} f(\boldsymbol{h})$ gleich dem Nullvektor $\boldsymbol{0}$ sein. Gleichsetzen liefert

$$2\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{h} - 2\boldsymbol{C}^T\boldsymbol{d} = \boldsymbol{0}. \tag{3.15}$$

Gleichung (3.15) kann durch einfaches Umformen bezüglich des Zielvektors $\boldsymbol{h}$ gelöst werden:

$$\begin{aligned} 2\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{h} - 2\boldsymbol{C}^T\boldsymbol{d} &= \boldsymbol{0}, \\ \boldsymbol{C}^T\boldsymbol{C}\boldsymbol{h} &= \boldsymbol{C}^T\boldsymbol{d}, \\ \boldsymbol{h} &= \left(\boldsymbol{C}^T\boldsymbol{C}\right)^{-1}\boldsymbol{C}^T\boldsymbol{d}, \\ \boldsymbol{h} &= \boldsymbol{C}^+\boldsymbol{d}, \end{aligned} \tag{3.16}$$

wobei $\boldsymbol{C}^+$ die Moore-Penrose Pseudoinverse der Matrix $\boldsymbol{C}$ ist [134]. Über die Moore-Penrose Pseudoinverse der Systemmatrix $\boldsymbol{C}$ wird ein lineares Gleichungssystem der Form

$$\boldsymbol{C}\boldsymbol{h} = \boldsymbol{d} \tag{3.17}$$

im Sinne des kleinsten quadratischen Fehlers bezüglich des Vektors $\boldsymbol{h}$ gelöst [134].

3.2.1.2 Optimierung des Systemabstandes

Der Filterentwurf mit gegebenem Zielsystem im Sinne des kleinsten quadratischen Fehlers muss nicht notwendigerweise mit Hilfe von statistischen Formulierungen erfolgen. Eine weitere Möglichkeit zur Herleitung ist es, direkt den euklidischen Abstand zwischen der globalen Impulsantwort $\boldsymbol{g} = \boldsymbol{C}\boldsymbol{h}$ und dem gewünschten Zielsystem $\boldsymbol{d}$ zu minimieren [30, 47, 89, 130]; $\boldsymbol{C}$ ist hierbei die Faltungsmatrix der Raumimpulsantwort $c(n)$, $\boldsymbol{h}$ und $\boldsymbol{d}$ sind wie in (3.4) definiert. Die Motivation für diese Herangehensweise ist direkt aus Gleichung (3.2) ersichtlich: Die Energie des Fehlersignal $e(n)$ wird minimal, wenn die Energie des Systemabstandes $\hat{e}(n) = g(n) - d(n)$ minimal ist.

Das sich aus dieser Motivation ergebende Optimierungsproblem lautet

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad : \quad f(\boldsymbol{h})\,, \quad f(\boldsymbol{h}) = \sum_{n=0}^{L_g-1} \left(g(n) - d(n)\right)^2 = \|\boldsymbol{C}\boldsymbol{h} - \boldsymbol{d}\|_2^2\,, \tag{3.18}$$

es soll also ein lineares Gleichungssystem der Form

$$\boldsymbol{C}\boldsymbol{h} - \boldsymbol{d} = \boldsymbol{0} \Longleftrightarrow \boldsymbol{C}\boldsymbol{h} = \boldsymbol{d} \tag{3.19}$$

im Sinne des kleinsten quadratischen Fehlers gelöst werden.

Mit

$$\|\boldsymbol{C}\boldsymbol{h} - \boldsymbol{d}\|_2^2 = (\boldsymbol{C}\boldsymbol{h} - \boldsymbol{d})^T(\boldsymbol{C}\boldsymbol{h} - \boldsymbol{d}) \tag{3.20}$$

kann die Kostenfunktion letztendlich als

$$f(\boldsymbol{h}) = \boldsymbol{h}^T\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{h} - 2\boldsymbol{h}^T\boldsymbol{C}^T\boldsymbol{d} + \boldsymbol{d}^T\boldsymbol{d} \tag{3.21}$$

geschrieben werden, was äquivalent zu (3.12) ist. Die Lösung zu (3.18) ist somit über die Moore-Penrose Pseudoinverse [134] der Systemmatrix $\boldsymbol{C}$ als

$$\boldsymbol{h} = \boldsymbol{C}^+\boldsymbol{d} \tag{3.22}$$

in geschlossener Form gegeben.

Beispiele An dieser Stelle wird anhand eines Beispiels die Entzerrung nach der Methode des kleinsten quadratischen Fehlers dargestellt. Hierzu wurde nach der Spiegelmethode (siehe Abschnitt 2.1.3.1 sowie [2]) eine Raumimpulsantwort $c(n)$ der Länge $L_c = 1000$ in einem Raum der Größe $[6.4 \times 4 \times 5]\ \mathrm{m}^3$ simuliert. Die Nachhallzeit wur-

de auf $T_{60} = 0.3$ Sekunden festgelegt und die Abtastrate wurde auf $f_s = 8000\,\mathrm{Hz}$ festgelegt.

Die simulierte Raumimpulsantwort auf einer logarithmischen Skala und der zugehörige Betragsfrequenzgang sind in Abbildung 3.3 dargestellt.

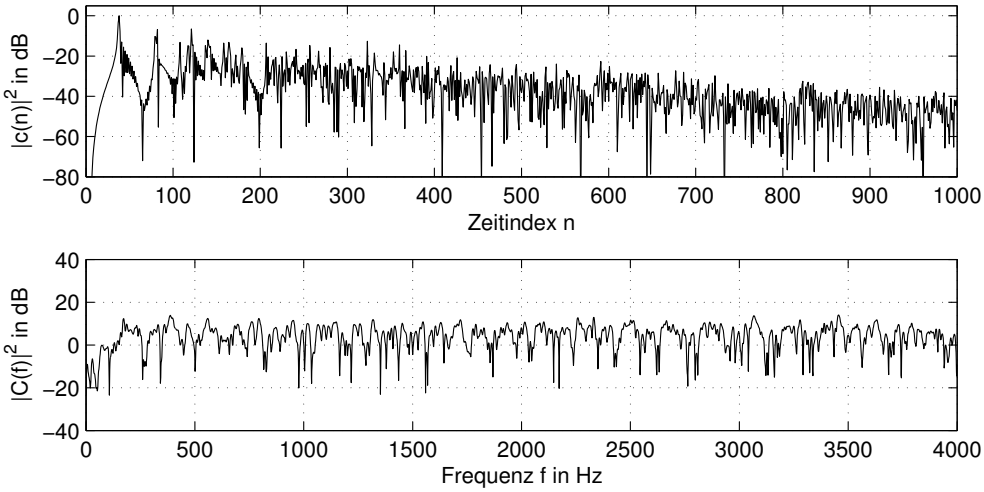

Abbildung 3.3: Darstellung einer simulierten Raumimpulsantwort im Zeit- (oben) und Frequenzbereich (unten).

Für die simulierte Raumimpulsantwort wird ein Entzerrer nach dem in diesem Abschnitt vorgestellten Verfahren über die Moore-Penrose Pseudoinverse [134] entworfen (3.22). Als Zielsystem $d_1(n)$ wurde ein verzögerter Einheitsimpuls $\delta(n - n_0)$ gewählt, wobei n_0 der Zeitindex der Direktschallkomponente der simulierten Raumimpulsantwort ist. Der Entzerrer $h_1(n)$ wurde mit einer Länge von $L_h = 1000$ Koeffizienten entworfen. Der so entworfene Entzerrer $h(n)$ wird mit der Raumimpulsantwort gefaltet, um die globale Impulsantwort $g_1(n) = h_1(n) * c(n)$ zu erzeugen.

Die globale Impulsantwort $g_1(n)$ auf einer logarithmischen Skala und der zugehörige Frequenzgang sind in Abbildung 3.4 dargestellt.

In einem zweiten Beispiel wird als gewünschtes Zielsystem ein verzögerter Einheitspuls verwendet, der um zusätzliche 10 Zeiteinheiten verzögert wird, also $d_2(n) = \delta(n - n_0 - 10)$. Der Entzerrer $h_2(n)$ wurde mit einer Länge von $L_h = 1000$ Koeffizienten entworfen. Das Ergebnis der Entzerrung $g_2(n) = h_2(n) * c(n)$ ist in Abbildung Abbildung 3.5 dargestellt.

Aus dem Vergleich der Abbildungen 3.4 und 3.5 ist ersichtlich, dass die Gesamtsysteme $g_1(n)$ und $g_2(n)$ unterschiedliche spektrale Eigenschaften aufweisen, obwohl

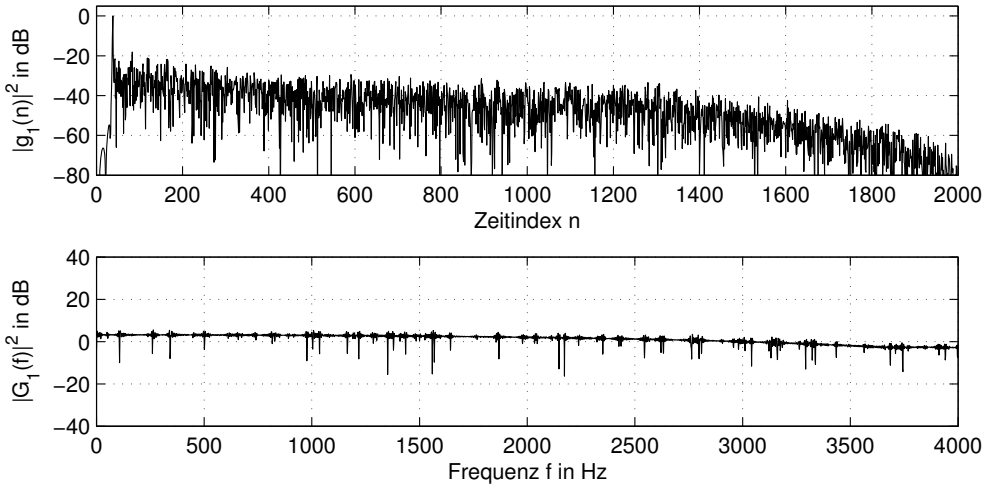

Abbildung 3.4: Darstellung der entzerrten Raumimpulsantwort im Zeit- (oben) und Frequenzbereich (unten). Als Zielsystem $d_1(n)$ wurde ein verzögerter Einheitsimpuls verwendet; die Verzögerung entspricht der Verzögerung der Direktschallkomponente der Raumimpulsantwort.

in beiden Fällen lediglich ein (unterschiedlich stark) verzögerter Einheitsimpuls mit flachem Betragsspektrum als zu approximierendes Zielsystem vorgegeben wurde.

3.2.1.3 Optimierung des gewichteten Systemabstandes

Zur Generalisierung der Minimierung des euklidischen Systemabstandes (3.18) wird ein Gewichtsfenster $w(n)$ eingeführt, welches den Fehler über die Zeitkoeffizienten bewertet. Entsprechende Verfahren wurden beispielsweise in [40] im Zusammenhang mit der automatischen Spracherkennung untersucht. Das zu lösende Optimierungsproblem lautet

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad : \quad \sum_{n=0}^{L_g-1} \left(w(n)(g(n)-d(n))\right)^2 = \|\boldsymbol{W}(\boldsymbol{C}\boldsymbol{h}-\boldsymbol{d})\|_2^2 \,, \tag{3.23}$$

wobei $\boldsymbol{W} = \mathrm{diag}\{\boldsymbol{w}\}$ mit

$$\boldsymbol{w} = [w(0)\,, w(1)\,, \ldots, w(L_g-1)]^T \tag{3.24}$$

gegeben ist.

Die Folge $w(n)$ enthält ein beliebiges Gewichtsfenster zur Bewertung des Fehlers. Für $w(n) = 1$, $0 \leq n \leq L_g - 1$, gilt $\boldsymbol{W} = \boldsymbol{I}$; in diesem Fall ist (3.23) äquivalent zu (3.18).

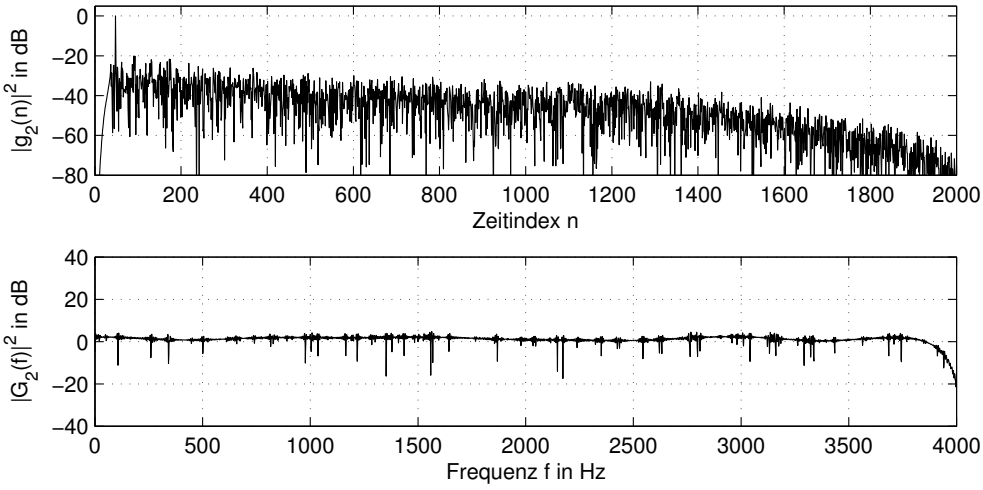

Abbildung 3.5: Darstellung der entzerrten Raumimpulsantwort im Zeit- (oben) und Frequenzbereich (unten). Als Zielsystem $d_2(n)$ wurde ein verzögerter Einheitsimpuls verwendet; die Verzögerung entspricht der Verzögerung der Direktschallkomponente der Raumimpulsantwort plus weiteren 10 Zeiteinheiten.

Durch das Gewichtsfenster $w(n)$ kann definiert werden, wie stark der Fehler zwischen der globalen Impulsantwort und dem gewünschten Zielsystem über die Zeit bewertet werden soll. Eine Motivation für die Einführung einer Gewichtsfunktion ist, dass eine *Formumg* der globalen Impulsantwort erreicht werden soll [40, 169]. Durch eine entsprechend *höhere* Bewertung der zeitlich späteren Teile der globalen Impulsantwort kann beispielsweise der späte Nachhall vermindert werden.

Eine andere Motivation ist die Definition sogenannter *don't care* Regionen; also Bereiche, in denen keine Anforderungen an die Zeitkoeffizienten der globalen Impulsantwort gestellt werden (in einigen Publikationen entsprechend als *relaxed methods* bezeichnet, zum Beispiel [101, 170]). Hierdurch können beispielsweise zugunsten einer *besseren* Formung im späten Teil der globalen Impulsantwort Freiheiten im frühen Teil der globalen Impulsantwort erlaubt werden [94, 95, 101, 170, 171].

Die Lösung zu (3.23) ist in geschlossener Form durch die Moore-Penrose Pseudoinverse [134] gegeben. Die Herleitung erfolgt analog zu der Lösung von (3.18). Auf die Herleitung im Detail soll an dieser Stelle verzichtet werden. Die Lösung von (3.23) lautet

$$\boldsymbol{h} = \underbrace{\left(\boldsymbol{C}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{C}\right)^{-1}\boldsymbol{C}^T\boldsymbol{W}^T}_{(\boldsymbol{W}\boldsymbol{C})^+}\boldsymbol{W}\boldsymbol{d} = (\boldsymbol{W}\boldsymbol{C})^+\boldsymbol{W}\boldsymbol{d}. \tag{3.25}$$

Beispiel Zur Demonstration der in diesem Abschnitt erklärten Methode zum Filterentwurf wird die bereits in Abschnitt 3.2.1 verwendete und in Abbildung 3.3 dargestellte simulierte Raumimpulsantwort verwendet.

Als Zielsystem wurde ein verzögerter Einheitsimpuls angesetzt, $d(n) = \delta(n - n_0)$, wobei n_0 der Verzögerung der Direktschallkomponente in der zu entzerrenden Raumimpulsantwort $c(n)$ entspricht. Der Entzerrer wurde mit einer Länge von $L_h = 1000$ Koeffizienten entworfen.

Für den Filterentwurf wurde ein Rechteckfenster zur Bewertung der Zeitkoeffizienten gewählt, welches eine *don't care* Region für die $50\,\mathrm{ms}$ ab der Direktschallkomponente definiert:

$$w(n) = \begin{cases} 0, & n_0 > n > n_0 + f_s\,0.05\,\mathrm{s} \\ 1, & \text{sonst.} \end{cases} \tag{3.26}$$

Das durch (3.26) definierte Gewichtsfenster ist in Abbildung 3.6 dargestellt.

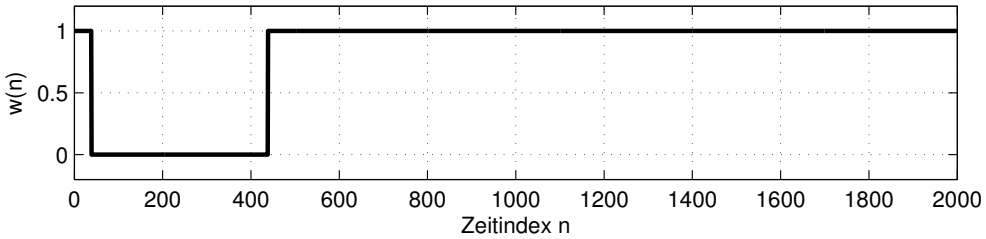

Abbildung 3.6: Darstellung des durch (3.26) definierten Gewichtsfensters. Das Fenster $w(n)$ definiert eine *don't care* Region, die die Fehlerenergie in den $50\,\mathrm{ms}$ ab dem Direktschall nicht in die Zielfunktion (3.23) einfließen lässt.

Das Ergebnis der Entzerrung bei Verwendung von $w(n)$ als Gewichtsfenster und (3.25) für den Filterentwurf ist in Abbildung 3.7 dargestellt.

3.2.2 Optimierung der Maximumsnorm

Das an dieser Stelle vorgestellte Verfahren von Terence Betlehem et al. [11] basiert auf dem von Mertins et al. [120] vorgeschlagenen Verfahren zur Formung von Raumimpulsantworten durch die Optimierung der Maximumsnorm.

Das von Betlehem et al. erarbeitete Verfahren wird in [11] zur Übersprechkompensation (siehe Abschnitt 3.5) angewendet. Aufgrund der Eigenschaft der verwendeten Zielfunktion („Approximation eines gewünschten Zielsystems") soll die Methode jedoch an dieser Stelle vorgestellt werden.

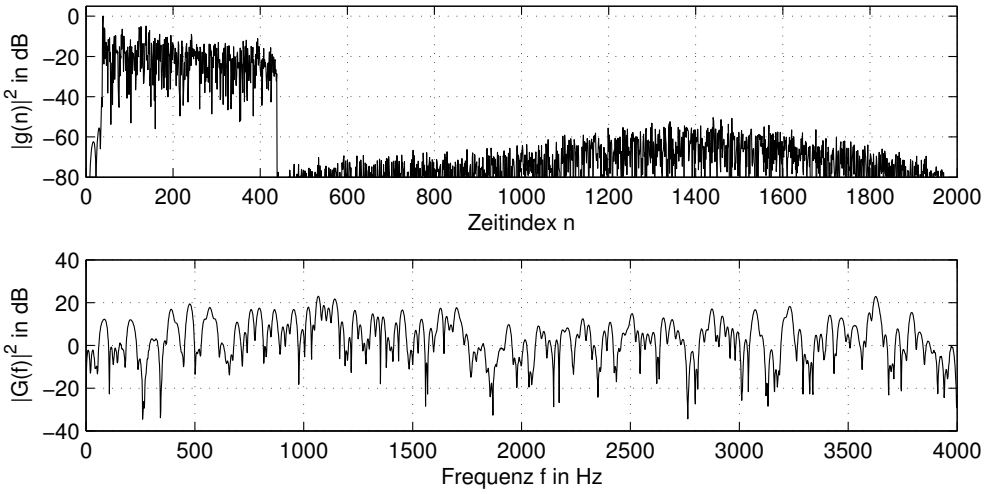

Abbildung 3.7: Darstellung der globalen Impulsantwort im Zeit- (oben) und Frequenzbereich (unten) bei Optimierung des gewichteten quadratischen Fehlers (3.23). Als Zielsystem $d(n)$ wurde ein verzögerter Einheitsimpuls angesetzt. Das verwendete Gewichtsfenster $w(n)$ ist in (3.26) definiert.

Die von Betlehem et al. vorgeschlagene Zielfunktion ist die Überführung von (3.23) von einer quadratischen Norm auf die Maximumsnorm. Die Motivation hierfür ist, dass die globale Impulsantwort keinen späten Anstieg, wie beispielsweise in Abbildung 3.7 zu sehen ist, aufweisen soll. Die Zielfunktion aus [11] lautet

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad : \quad \|\boldsymbol{W}(\boldsymbol{C}\boldsymbol{h} - \boldsymbol{d})\|_\infty \,. \tag{3.27}$$

Wie in (3.23) bezeichnet $\boldsymbol{h}$ das zu entwerfende Filter. Die Matrix $\boldsymbol{W} = \operatorname{diag}\{\boldsymbol{w}\}$ mit $\boldsymbol{w} = [w(n)\,, \ldots, w(L_g - 1)]^T$ gewichtet den Systemabstand zwischen der globalen Impulsantwort und dem gewünschten Zielsystem über die Zeit; $\boldsymbol{d}$ enthält das gewünschte Zielsystem $d(n)$ und $\boldsymbol{C}$ bezeichnet die Faltungsmatrix der Raumimpulsantwort $c(n)$.

Durch das Einführen einer Schlupfvariable τ kann das auf der Maximumsnorm basierende Optimierungsproblem (3.27) in Epigraph-Form [14] geschrieben werden:

$$\begin{aligned} \underset{\tau,\boldsymbol{h}}{\text{minimize}} \quad &: \quad \tau \\ \text{subject to} \quad &: \quad [\boldsymbol{W}(\boldsymbol{C}\boldsymbol{h} - \boldsymbol{d})]_n \leq \tau, \\ & \quad [\boldsymbol{W}(\boldsymbol{C}\boldsymbol{h} - \boldsymbol{d})]_n \geq -\tau, \quad n = 1, \ldots, L_g, \end{aligned} \tag{3.28}$$

wobei $[\boldsymbol{W}(\boldsymbol{C}\boldsymbol{h} - \boldsymbol{d})]_n$ den n-ten Eintrag des durch $\boldsymbol{W}(\boldsymbol{C}\boldsymbol{h} - \boldsymbol{d})$ gegebenen Vektors bezeichnet; L_g ist die Länge der globalen Impulsantwort $g(n)$.

Das in (3.28) gegebene Optimierungsproblem enthält lediglich lineare Ungleichungen; auch die Zielfunktion ist linear. Somit kann (3.28) durch Verfahren der linearen Programmierung gelöst werden [14].

Wird das Optimierungsproblem (3.28) in ein lineares Programm in Standardform überführt, so hat die Systemmatrix die Größe $(2L_g + 1) \times (L_h + 1)$.

Da Raumimpulsantworten und die zu entwerfenden Entzerrer typischerweise mehrere tausend Koeffizienten im Zeitbereich besitzen ist die Lösung von (3.28) üblicherweise sehr speicher- und rechenaufwendig [11]. Im Rahmen der numerischen Rechengenauigkeit kann die exakte Lösung eines linearen Optimierungsproblems gefunden werden [136].

Beispiel Zur Demonstration der in diesem Abschnitt vorgestellten Methode von Betlehem et al. [11] wird die in Abbildung 3.3 dargestellte, simulierte Raumimpulsantwort verwendet. Als Gewichtsfenster wurde ein Rechteckfenster (siehe Abbildung 3.6) verwendet.

Zum Entwurf der Entzerrer wurde (3.27) beziehungsweise (3.28) gelöst. Die globale Impulsantwort bei einem mit dem Rechteckfenster gewichteten Filterentwurf ist in Abbildung 3.8 dargestellt.

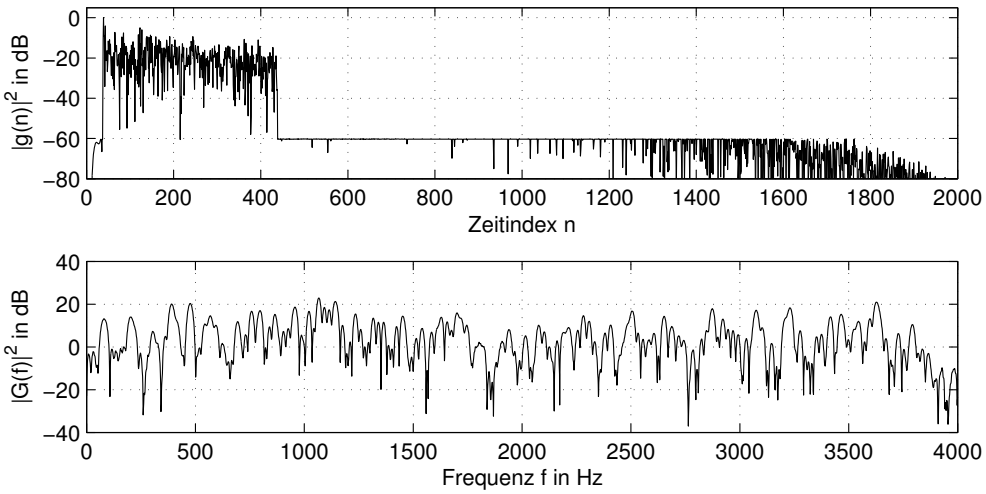

Abbildung 3.8: Darstellung der entzerrten Raumimpulsantwort im Zeit- (oben) und Frequenzbereich (unten) bei Optimierung der Maximumsnorm (3.27). Als Zielsystem $d(n)$ wurde ein verzögerter Einheitsimpuls angesetzt. Das verwendete Gewichtsfenster $w(n)$ ist in (3.26) definiert.

3.2.3 Zusammenfassung

In diesem Abschnitt wurde die Kompensation akustischer Kanäle durch Approximation eines vorgegebenen Zielsystems erklärt. Es wurde der Filterentwurf durch Optimierung einer quadratischen Norm und der Maximumsnorm vorgestellt.

Die Güte der Kanalkompensation ist abhängig von der Wahl des Zielsystems [96]. Durch die Vorgabe des Zielsystems lassen sich beispielsweise eine Verzögerung oder gewünschte spektrale Eigenschaften genau festlegen.

Alle Verfahren wurden mit Beispielen vorgestellt; auch der Einfluss der Verzögerung als ein freier Parameter des Zielsystems auf die Entzerrung wurde anhand eines Beispiels demonstriert.

Die Entscheidung über die Wahl des zu approximierenden Zielsystems muss vor dem eigentlichen Filterentwurf erfolgen, was ein essentieller Nachteil dieser Verfahren ist. Ein erster Versuch, die Parameter des *besten* Zielsystems aus der zu entzerrenden Raumimpulsantwort abzuleiten, wurde in [43] unternommen. Goetze et al. konnten in [43] zeigen, dass die ideale Verzögerung des Zielsystems mit der Zentrums-Zeit der Raumimpulsantwort (2.18) positiv korreliert ist.

3.3 Kanalkompensation ohne Vorgabe eines Zielsystems

Wie bereits im vorherigen Abschnitt angedeutet ist das größte Problem bei der Entzerrung durch die Approximation eines gewünschten Zielsystems die Wahl eben dieses Zielsystems. Die Güte der Entzerrung ist abhängig von der Wahl des Zielsystems [96]. Freie Parameter sind beispielsweise Bandpass-Charakteristika und Verzögerungen des Zielsystems.

In diesem Abschnitt werden Methoden zur Kanalkompensation ohne vorgegebenes Zielsystem vorgestellt. Diese Verfahren basieren im Allgemeinen darauf, dass die globale Impulsantwort in einen *gewünschten* und einen *ungewünschten* Teil aufgeteilt wird. Der Filterentwurf geschieht dann durch die Maximierung einer vom gewünschten Teil abhängigen Funktion bei gleichzeitiger Minimierung einer vom ungewünschten Teil abhängigen Funktion.

Auf diese Weise ist es möglich, lediglich die *Form* (Impulsantwortformung, englisch *impulse response reshaping*) oder die gewünschte effektive *Länge* (Impulsantwortverkürzung, englisch *channel shortening*) des Gesamtsystems als Optimalitätskriterium zu

formulieren. Entsprechende Verfahren kommen ursprünglich aus der Datenübertragung und werden dort zur Verkürzung der effektiven Länge der Übertragungskanäle benutzt [4, 108, 118, 163].

3.3.1 Quadratisches Optimalitätskriterium

Die Methode zur Reduzierung der effektiven Kanallänge wird auch als MSSNR (englisch *maximum shortening signal-to-noise ratio*) bezeichnet [4]. Die Methode wurde ursprünglich von Melsa et al. vorgeschlagen [118]. Kallinger et al. untersuchten in [85] die Anwendbarkeit dieses Verfahrens auf akustische Kanäle.

Für das MSSNR Verfahren wird durch ein Gewichtsfenster $w_u(n)$ ein *ungewünschter* (englisch *unwanted*) Teil der globalen Impulsantwort als

$$g_u(n) = w_u(n)(c(n) * h(n)) = w_u(n)\, g(n) \tag{3.29}$$

definiert, in dem die Energie minimiert werden soll. Gleichzeitig wird durch ein zweites Gewichtsfenster $w_d(n)$ ein *gewünschter* (englisch *desired*) Teil der globalen Impulsantwort als

$$g_d(n) = w_d(n)(c(n) * h(n)) = w_d(n)\, g(n) \tag{3.30}$$

definiert, in dem die Energie konstant bleiben oder maximiert werden soll [4, 85, 118].

Das für den Filterentwurf zu lösende Optimierungsproblem lautet

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad : \quad \sum_{n=0}^{L_g-1} g_u^2(n) \quad \text{s.t.} \quad \sum_{n=0}^{L_g-1} g_d^2(n) = 1. \tag{3.31}$$

Alternativ kann (3.31) auch als Maximierungsproblem formuliert werden:

$$\underset{\boldsymbol{h}}{\text{maximize}} \quad : \quad \sum_{n=0}^{L_g-1} g_d^2(n) \quad \text{s.t.} \quad \sum_{n=0}^{L_g-1} g_u^2(n) = 1, \tag{3.32}$$

wobei die Energie im gewünschten Teil der globalen Impulsantwort maximiert wird, während die Energie im ungewünschten Teil der globalen Impulsantwort konstant bleibt.

Mit

$$\boldsymbol{w}_d = [w_d(0)\,, w_d(1)\,, \ldots, w_d(L_g-1)]^T \tag{3.33}$$

und

$$\boldsymbol{w}_u = [w_u(0)\,, w_u(1)\,, \ldots, w_u(L_g-1)]^T \tag{3.34}$$

lässt sich der gewünschte Teil der globalen Impulsantwort als

$$\boldsymbol{g}_d = [g_d(0)\,, g_d(1)\,, \ldots, g_d(L_g-1)]^T = \mathrm{diag}\{\boldsymbol{w}_d\}\,\boldsymbol{C}\boldsymbol{h} \tag{3.35}$$

und der ungeschwünschte Teil analog als

$$\boldsymbol{g}_u = [g_u(0)\,, g_u(1)\,, \ldots, g_u(L_g-1)]^T = \mathrm{diag}\{\boldsymbol{w}_u\}\,\boldsymbol{C}\boldsymbol{h} \tag{3.36}$$

ausdrücken. Mit 3.35 und 3.36 kann die Energie im gewünschten Teil der globalen Impulsantwort als

$$\boldsymbol{g}_d^T\boldsymbol{g}_d = \boldsymbol{h}^T\underbrace{\boldsymbol{C}^T\mathrm{diag}\{\boldsymbol{w}_d\}^T\,\mathrm{diag}\{\boldsymbol{w}_d\}\,\boldsymbol{C}}_{\boldsymbol{B}}\boldsymbol{h} = \boldsymbol{h}^T\boldsymbol{B}\boldsymbol{h}, \tag{3.37}$$

beziehungsweise im ungewünschten Teil als

$$\boldsymbol{g}_u^T\boldsymbol{g}_u = \boldsymbol{h}^T\underbrace{\boldsymbol{C}^T\mathrm{diag}\{\boldsymbol{w}_u\}^T\,\mathrm{diag}\{\boldsymbol{w}_u\}\,\boldsymbol{C}}_{\boldsymbol{A}}\boldsymbol{h} = \boldsymbol{h}^T\boldsymbol{A}\boldsymbol{h}, \tag{3.38}$$

berechnet werden, wobei $\boldsymbol{C}$ die Faltungsmatrix der Raumimpulsantwort $c(n)$ bezeichnet.

In Vektor-Notation kann das zu lösende Optimierungsproblem (3.32) mit (3.37) und (3.38) als Maximierung des generalisierten Rayleigh Quotienten

$$\underset{\boldsymbol{h}}{\mathrm{maximize}} \quad : \quad \frac{\|\mathrm{diag}\{\boldsymbol{w}_d\}\,\boldsymbol{C}\boldsymbol{h}\|_2^2}{\|\mathrm{diag}\{\boldsymbol{w}_u\}\,\boldsymbol{C}\boldsymbol{h}\|_2^2} = \frac{\boldsymbol{h}^T\boldsymbol{B}\boldsymbol{h}}{\boldsymbol{h}^T\boldsymbol{A}\boldsymbol{h}} \tag{3.39}$$

ausgedrückt werden [95, 118]. Die Lösung von (3.39) ist durch das generalisierte Eigenwertproblem in geschlossener Form gegeben:

$$\boldsymbol{B}\boldsymbol{h}_{\text{opt}} = \lambda_{\text{max}}\boldsymbol{A}\boldsymbol{h}_{\text{opt}}, \tag{3.40}$$

wobei λ_{max} den größten Eigenwert und $\boldsymbol{h}_{\text{opt}}$ den zugehörigen Eigenvektor bezeichnet [85, 95, 118].

Des Weiteren wurde von Kallinger et al. [85] empfohlen, eine *Formung* der globalen Impulsantwort einer *Verkürzung* vorzuziehen. Hierdurch können die psychoakustischen Eigenschaften des menschlichen Hörapparates besser ausgenutzt werden. Während bei einer Verkürzung üblicherweise einfache Rechteckfenster verwendet werden [4, 85, 118] sind bei einer Formung beliebige Gewichtsfunktionen erlaubt; üblicherweise wird jedoch ein schnelles Abklingen der globalen Impulsantwort gefordert [85, 120].

Beispiel Zur Demonstration des MSSNR Verfahrens wird die in Abbildung 3.3 dargestellte Raumimpulsantwort verwendet. Für das Experiment werden die Gewichtsfenster so gewählt, dass, wie in [85] vorgeschlagen, das D_{50} Maß (2.13) maximiert wird. Das D_{50} Maß ist definiert als die Energie in den ersten 50 Millisekunden beginnend mit dem Direktschall im Verhältnis zur Gesamtenergie der globalen Impulsantwort [98]. Es wird ein Fenster $w_d(n)$ als

$$w_d(n) = \begin{cases} 1, & n_0 \leq n \leq n_0 + f_s\, 50\,\mathrm{ms} \\ 0, & \text{sonst} \end{cases} \tag{3.41}$$

definiert, wobei f_s die Abtastrate und n_0 die Position der Direktschallkomponente bezeichnet. Das Fenster für den ungewünschten Teil wird als

$$w_u(n) = 1 - w_d(n) \tag{3.42}$$

berechnet [85].

Die durch (3.41) und (3.42) definierten Fenster zur direkten Optimierung des D_{50} Maßes sind in Abbildung 3.9 dargestellt. Bei der Optimierung von (3.39) mit den in (3.41) und (3.42) definierten Fenstern spricht man von einer *Kanalverkürzung* [85].

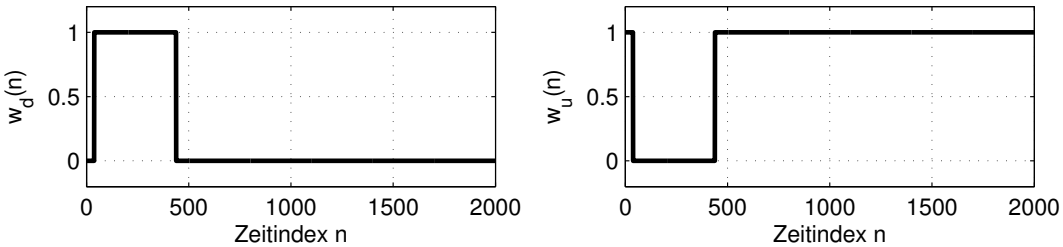

Abbildung 3.9: Gewichtsfenster nach [85] zur direkten Optimierung des D_{50} Maßes [98]. Das Fenster $w_d(n)$ (links) selektiert die ersten 50 Millisekunden, beginnend mit der Direktschallkomponente, der globalen Impulsantwort. Das Fenster für den ungewünschten Teil (rechts) wird als $w_u(n) = 1 - w_d(n)$ bestimmt.

Werden $w_d(n)$ und $w_u(n)$ als Gewichtsfenster für den Filterentwurf durch Lösung von (3.39) verwendet, so wird das in Abbildung 3.10 dargestellte Ergebnis erreicht. Das D_{50} Maß [98] konnte durch die Entzerrung von 0.83 auf 1.00 gesteigert werden.

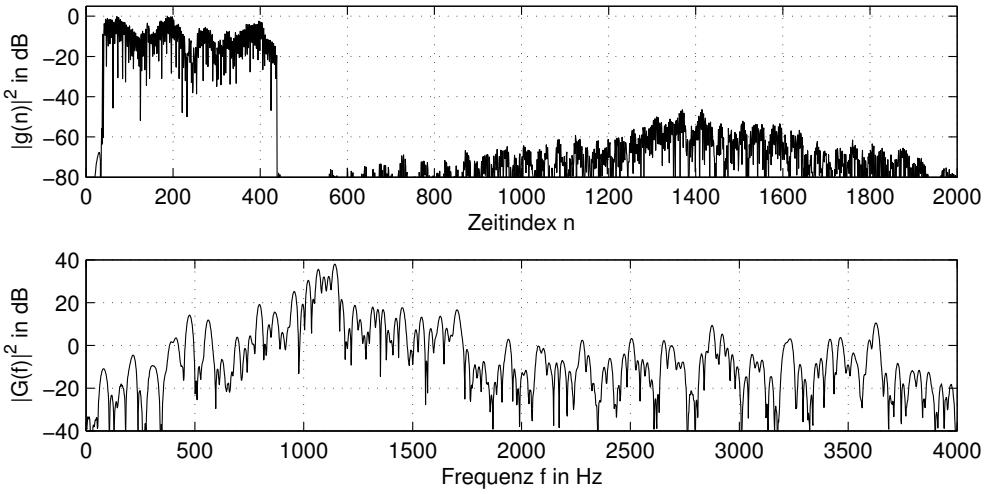

Abbildung 3.10: Darstellung der entzerrten Raumimpulsantwort im Zeit- (oben) und Frequenzbereich (unten). Durch die Wahl der Gewichtsfenster konnte, wie in [85] vorgeschlagen, das D_{50} Maß maximiert werden. Das D_{50} Maß wurde durch die Impulsformung nach dem MSSNR Verfahren von 0.83 auf 1.00 erhöht.

3.3.2 Spektrale Aspekte

Wird das zu approximierende System direkt gewählt (siehe Abschnitt 3.2), so haben die spektralen Eigenschaften des Zielsystems direkten Einfluss auf das globale System.

Erfolgt der Filterentwurf ohne Vorgabe eines Zielsystems, so können spektrale Störungen im Gesamtsystem auftreten [81, 85]. Die üblicherweise verwendeten quadratischen Zielfunktionen [85] zur Kanalverkürzung optimieren lediglich die Zeitbereichsdarstellung des Gesamtsystems und lassen spektrale Anforderungen vollkommen unbeachtet [81].

Die durch das MSSNR Verfahren induzierten spektralen Störungen sind beispielsweise in dem in Abbildung 3.10 dargestellten Betragsfrequenzgang zu erkennen.

Gegenstand dieses Abschnitts ist die Kompensation der spektralen Störungen. Hierzu wird das Verfahren von Kallinger et al. aus [85] vorgestellt.

Postfilter Methode In [85] wurde die Anwendbarkeit von Verfahren zur Kanalverkürzung auf akustische Kanäle untersucht; dieses Verfahren wurde in Abschnitt 3.3.1 vorgestellt.

Um spektrale Färbungen des Audiosignals zu vermeiden, fordern Kallinger et al. einen flachen Betragsfrequenzgang des Gesamtsystems [85]. Hierzu empfehlen Kallinger et al. die Anwendung eines Postfilters.

Für das Gesamtsystem wird ein kurzes Ein-Schritt Prädiktorfehlerfilter [119] entworfen und auf die geformte Impulsantwort angewendet. Die nachgefilterte globale Impulsantwort $\hat{g}(n)$ ist nach dem Verfahren aus [85] durch

$$\hat{g}(n) = \underbrace{c(n) * h(n)}_{g(n)} * a(n) \tag{3.43}$$

gegeben, wobei $c(n)$ die Raumimpulsantwort, $h(n)$ den Entzerrer und $a(n)$ den Ein-Schritt Prädiktorfehlerfilter bezeichnet. Das Prädiktorfilter $\hat{a}(n)$ für das Prädiktorfehlerfilter $a(n)$ wurde für das Faltungsprodukt aus Raumimpulsantwort und Entzerrer, wie in Abbildung 3.11 dargestellt, durch Minimierung der Fehlerenergie $\mathrm{E}\{e^2(n)\}$ entworfen.

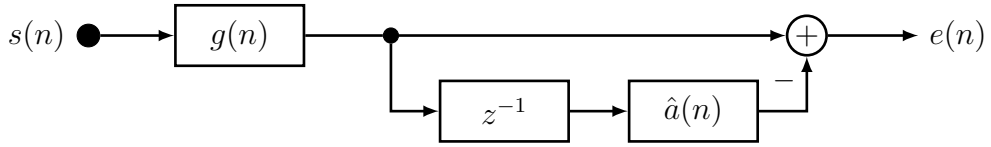

Abbildung 3.11: Signalflussgraph für den Entwurf des linearen Prädiktorfilters für die Postfilter-Methode nach [85].

Beispiel Zur Demonstration der Postfilter-Methode nach Kallinger et al. [85] wird die verkürzte Impulsantwort aus dem Beispiel aus Abschnitt 3.3.1 (Abbildung 3.10) verwendet.

Für die globale Impulsantwort wird ein lineares Prädiktorfehlerfilter $a(n)$ der Länge $L_a = 40$ berechnet, um die spektralen Störungen zu kompensieren und einen flachen Betragsfrequenzgang zu erzeugen. Durch die zusätzliche Nachfilterung verringert sich der Wert des D_{50} Maßes von 1.00 auf 0.98. Die globale Impulsantwort nach der zusätzlichen Filterung ist in Abbildung 3.12 dargestellt.

3.3.3 Maximumsnormbasierte Formung

Ein Nachteil der in Abschnitt 3.3.1 beschriebenen Methode zur Impulsantwortformung oder Verkürzung sind die spektralen Artefakte, die dem Gesamtsystem durch die Entzerrung aufgeprägt werden [85]. In diesem Abschnitt werden zwei Verfahren zur Impulsantwortformung auf Basis der Maximumsnorm vorgestellt.

Motiviert werden die Verfahren durch die Überlegung, dass die Maximierung der Maximumsnorm im gewünschten Teil der globalen Impulsantwort zur Ausprägung eines einzelnen dominanten Impulses und damit zu einem flachen Betragsfrequenzgang führt

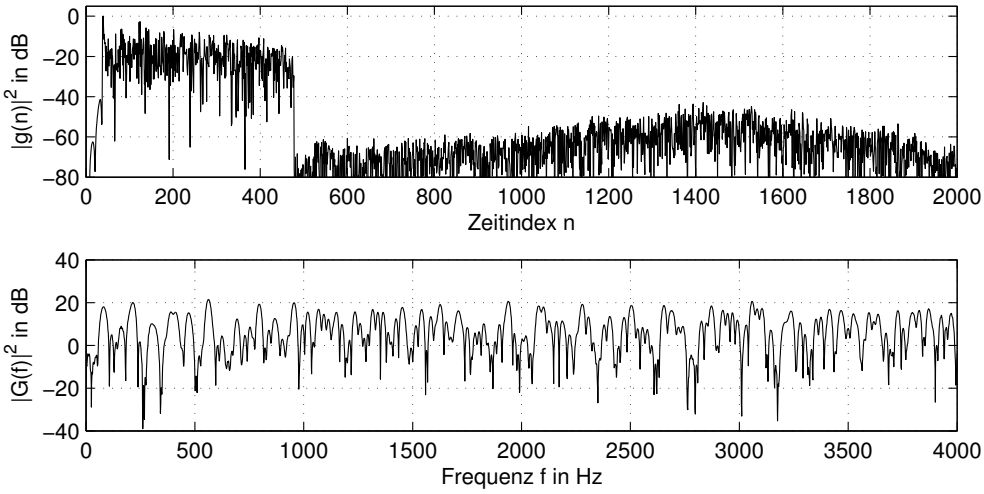

Abbildung 3.12: Die globale Impulsantwort bei Anwendung des MSSNR Verfahrens mit einer Nachfilterung nach [85] im Zeit- (oben) und Frequenzbereich (unten). Durch die Wahl der Gewichtsfenster wurde, wie in [85] vorgeschlagen, das D_{50} Maß maximiert. Durch die Nachfilterung verringert sich das D_{50} Maß von 1.00 (Abbildung 3.10) auf 0.98.

[115, 120]. Die gewünschten spektralen Eigenschaften sollen also implizit über das im Zeitbereich formulierte Optimalitätskriterium herbeigeführt werden.

3.3.3.1 Impulsantwortformung mit einer maximumsnormbasierten Nebenbedingung

Von Mei et al. wurde in [115] ein Optimalitätskriterium vorgeschlagen, welches die Energie im ungewünschten Teil der globalen Impulsantwort minimiert, während die Maximumsnorm des gewünschten Teils der globalen Impulsantwort konstant bleiben soll. Das in [115] vorgeschlagene Optimierungsproblem lautet

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad : \quad \boldsymbol{g}_u^T \boldsymbol{g}_u \quad \text{s.t.} \quad \|\boldsymbol{g}_d\|_\infty = 1, \tag{3.44}$$

wobei $\boldsymbol{g}_u = \mathrm{diag}\{\boldsymbol{w}_u\}\,\boldsymbol{C}\boldsymbol{h}$ den ungewünschten Teil der globalen Impulsantwort und $\boldsymbol{g}_d = \mathrm{diag}\{\boldsymbol{w}_d\}\,\boldsymbol{C}\boldsymbol{h}$ den gewünschten Teil der globalen Impulsantwort enthält (siehe Gleichungen (3.30) und (3.29)). Die Matrix $\boldsymbol{C}$ bezeichnet die Faltungsmatrix der Raumimpulsantwort und die Vektoren $\boldsymbol{w}_d$ und $\boldsymbol{w}_u$ enthalten die Gewichtsfenster für den gewünschten sowie den ungewünschten Teil der globalen Impulsantwort.

Zur Lösung von (3.44) wird eine Matrix $\boldsymbol{A}$ als

$$\boldsymbol{A} = \boldsymbol{C}^T \mathrm{diag}\{\boldsymbol{w}_u\}^T \, \mathrm{diag}\{\boldsymbol{w}_u\} \, \boldsymbol{C} \tag{3.45}$$

berechnet. Des Weiteren wird gefordert, dass die Formung zu keiner Verzögerung in dem Gesamtsystem führt und sowohl die Raumimpulsantwort $c(n)$ als auch die globale Impulsantwort $g(n)$ auf einen Maximalwert von eins normalisiert sind. Bezeichne n_{max} den diskreten Zeitindex der Direktschallkomponente in der Raumimpulsantwort $c(n)$, so gilt mit diesen Annahmen

$$|c(n_{\text{max}})| = c(n_{\text{max}}) = |g(n_{\text{max}})| = g(n_{\text{max}}) = \|\boldsymbol{g}\|_\infty = 1. \tag{3.46}$$

Der Vektor

$$\boldsymbol{c}_{\text{n-max}} = [1, c(n_{\text{max}} - 1), \ldots, c(0), 0, \ldots, 0]^T \tag{3.47}$$

ist die Transponierte der n_{max}-ten Zeile der Faltungsmatrix $\boldsymbol{C}$.

Mit (3.47), (3.45) und (3.46) kann das Optimierungsproblem (3.44) durch Anwendung der Lagrange-Methode als

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{c}_{\text{n-max}} \\ \boldsymbol{c}_{\text{n-max}}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{h} \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \tag{3.48}$$

mit dem Lagrange-Multiplikator λ formuliert werden.

Die Lösung zu (3.44) ist über (3.48) durch

$$\boldsymbol{h} = \frac{\boldsymbol{A}^{-1}\boldsymbol{c}_{\text{n-max}}}{\boldsymbol{c}_{\text{n-max}}^T \boldsymbol{A}^{-1}\boldsymbol{c}_{\text{n-max}}} \tag{3.49}$$

in geschlossener Form gegeben [115].

Beispiel Zur Demonstration des Verfahrens nach [115] wurde eine simulierte Raumimpulsantwort (siehe Abbildung 3.3) verwendet. Der Entzerrer wurde mit einer Länge von $L_h = 1000$ Zeitkoeffizienten entworfen. Für das Beispiel wurde das Gewichtsfenster $w_u(n)$ wie in [85] vorgeschlagen dem D_{50} Maß [98] entsprechend gewählt (siehe Gleichung (3.42)). Das Ergebnis der Entzerrung ist in Abbildung 3.13 dargestellt.

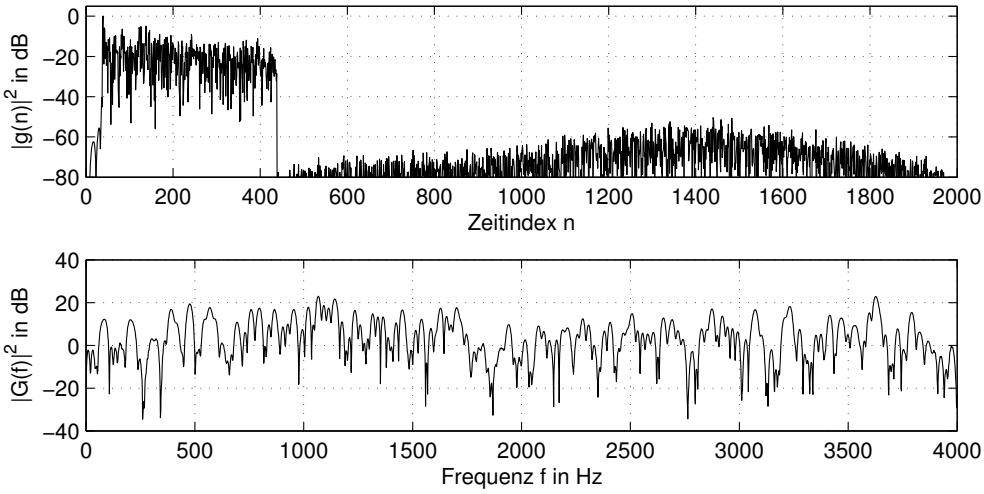

Abbildung 3.13: Darstellung der nach dem Verfahren aus [115] entzerrten Raumimpulsantwort im Zeit- (oben) und Frequenzbereich (unten). Die Gewichtsfenster wurden, wie in [85] vorgeschlagen, als Rechteckfenster gewählt.

3.3.3.2 Formung durch Optimierung der Maximumsnorm

Wie in den Beispielen in diesem Abschnitt zu sehen, führt die Optimierung eines quadratischen Optimalitätskriteriums üblicherweise zu einem diffusen späten Nachhall, exemplarisch zu erkennen in den Abbildungen 3.7, 3.10 und 3.13 jeweils im Bereich $1200 \leq n \leq 1900$. Die Zeitkoeffizienten sind betragsmäßig *klein* und minimieren entsprechend die Zielfunktion, jedoch sind sie als störender Nachhall wahrnehmbar [116, 120]. Des Weiteren ist die Formung bei Optimierung eines quadratischen Kriteriums *ungleichmäßiger* als bei Optimierung eines auf der Maximumsnorm basierenden Kriteriums; ein Vergleich der Beispiele in den Abschnitte 3.2.1 und 3.2.2 verdeutlicht dies.

Das von Mei et al. in [116] empfohlene Optimierungsproblem kann als die Überführung der quadratischen Norm von (3.39) auf die Maximumsnorm verstanden werden.

Das Optimierungsproblem aus [116] lautet

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad : \quad f(\boldsymbol{h}) \,, \quad f(\boldsymbol{h}) = \log\left(\frac{f_u(\boldsymbol{h})}{f_d(\boldsymbol{h})}\right) , \tag{3.50}$$

mit

$$f_d(\boldsymbol{h}) = \|\boldsymbol{g}_d\|_\infty = \|\text{diag}\{\boldsymbol{w}_d\}\, \boldsymbol{C}\boldsymbol{h}\|_\infty \tag{3.51}$$

und

$$f_u(\boldsymbol{h}) = \|\boldsymbol{g}_u\|_\infty = \|\mathrm{diag}\{\boldsymbol{w}_u\}\,\boldsymbol{C}\boldsymbol{h}\|_\infty\,, \tag{3.52}$$

wobei $\boldsymbol{C}$ die Faltungsmatrix der Raumimpulsantwort ist und die Vektoren $\boldsymbol{w}_d$ und $\boldsymbol{w}_u$ die Gewichtsfenster $w_d(n)$ und $w_u(n)$ für den gewünschten und den ungewünschten Teil der globalen Impulsantwort enthalten.

Bedeutung der Gewichtsfenster Die Motivation für die Maximierung der Maximumsnorm im gewünschten Teil der globalen Impulsantwort ist die Überlegung, dass sich ein einzelner dominanter Zeitkoeffizient ausbildet und das Gesamtsystem somit einen flachen Betragsfrequenzgang aufweist. Die Minimierung der Maximumsnorm des ungewünschten Teils der globalen Impulsantwort ist an die Equiripple-Entwurfsmethode für FIR-Filter [132] angelehnt. Durch die sukzessive Minimierung des jeweils betragsmäßig maximalen Zeitkoeffizienten im ungewünschten Teil $g_u(n)$ (3.29) der globalen Impulsantwort konvergieren alle Zeitkoeffizienten gegen den gleichen Betrag, was zu einer sehr gleichmäßigen Formung führt [79].

Bei vollständiger Optimierung von (3.50) ist der Maximalwert von $|g_u(n)|$ gleich einer Konstante τ; wobei natürlich mehrere Koeffizienten $|g_u(n)|$ den Wert τ annehmen können. Damit gilt

$$|g_u(n)| \leq \tau / w_u(n) \quad \text{für } n \geq n_1, \tag{3.53}$$

wobei gilt $w_u(n) > 0$ für $n \geq n_1$; n_1 ist ein beliebiger Zeitindex. Anders ausgedrückt bedeutet dies, dass der Betrag des ungewünschte Teils der globalen Impulsantwort $|g_u(n)|$ durch das Reziproke des Gewichtsfensters $w_u(n)$ mal einem Faktor τ begrenzt ist.

Iterative Optimierung der Zielfunktion Die Lösung zu (3.50) kann nicht in geschlossener Form angegeben werden. Zur Lösung von (3.50) wird ein Gradientenabstiegsverfahren [14] angewendet, wobei die Logarithmus-Operation in (3.50) zu einer kompakteren Darstellung der Gradienten führt [116, 120].

Die Lernregel lautet

$$\boldsymbol{h}^{l+1} = \boldsymbol{h}^l - \mu\left(\frac{1}{f_u(\boldsymbol{h}^l)}\nabla_{\boldsymbol{h}} f_u(\boldsymbol{h}^l) - \frac{1}{f_d(\boldsymbol{h}^l)}\nabla_{\boldsymbol{h}} f_d(\boldsymbol{h}^l)\right), \tag{3.54}$$

wobei μ die feste Schrittweite des Gradientenabstiegsverfahrens bezeichnet; l bezeichnet die aktuelle Iteration des Abstiegsverfahrens.

Bezeichne n_d den diskreten Zeitindex des betragsmäßig größten Eintrages des gewünschten Anteils und n_u den diskreten Zeitindex des betragsmäßig größten Eintrages des ungewünschten Anteils der globalen Impulsantwort. Es gilt dann $f_d(\boldsymbol{h}) = \|\boldsymbol{g}_d\|_\infty = |g_d(n_d)|$ und $f_u(\boldsymbol{h}) = \|\boldsymbol{g}_u\|_\infty = |g_u(n_u)|$.

Die Gradienten für $f_d(\boldsymbol{h})$ und $f_u(\boldsymbol{h})$ bezüglich der Filterkoeffizienten $h(n)$ sind durch

$$\nabla_{\boldsymbol{h}} f_d(\boldsymbol{h}) = \operatorname{sign}\{g_d(n_d)\}\, w_d(n_d)\, \boldsymbol{C}_{n_d}^T \tag{3.55}$$

und

$$\nabla_{\boldsymbol{h}} f_u(\boldsymbol{h}) = \operatorname{sign}\{g_u(n_u)\}\, w_u(n_u)\, \boldsymbol{C}_{n_u}^T \tag{3.56}$$

gegeben, wobei $\boldsymbol{C}_n$ die n-te Zeile der Faltungsmatrix $\boldsymbol{C}$ bezeichnet.

Beispiel Zur Demonstration des Verfahrens nach [116] wurde eine simulierte Raumimpulsantwort (siehe Abbildung 3.3) verwendet. Der Entzerrer wurde mit einer Länge von $L_h = 1000$ Zeitkoeffizienten entworfen. Als Schrittweite wurde $\mu = 10^{-7}$ gewählt und es wurden $2{\cdot}10^7$ Iterationen des Gradientenabstiegsverfahrens durchgeführt. Für das Beispiel wurden die Gewichtsfenster $w_d(n)$ und $w_u(n)$, wie in [85] vorgeschlagen, dem D_{50}-Maß [98] entsprechend gewählt (siehe Gleichung (3.41) und (3.41)). Das Ergebnis der Formung ist in Abbildung 3.14 dargestellt.

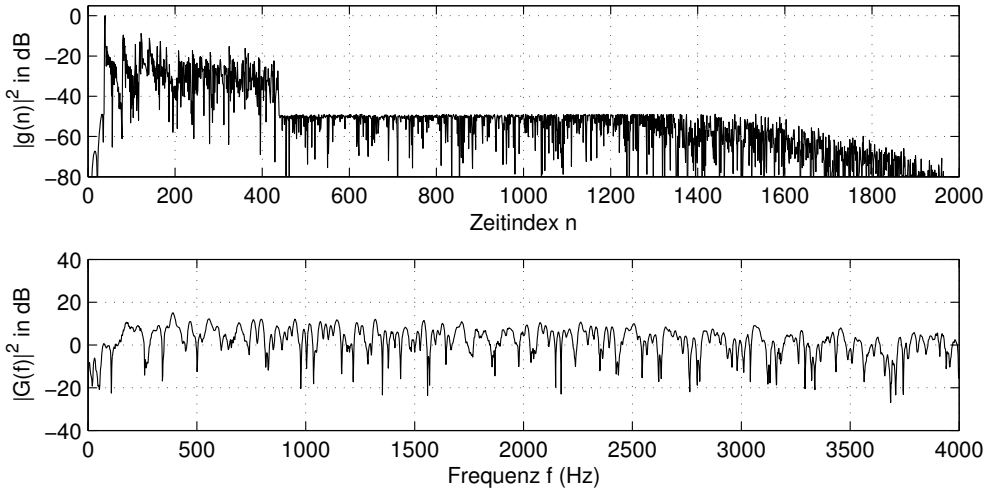

Abbildung 3.14: Darstellung der nach dem Verfahren aus [116] entzerrten Raumimpulsantwort im Zeit- (oben) und Frequenzbereich (unten). Die Gewichtsfenster wurden, wie in [85] vorgeschlagen, als Rechteckfenster gewählt.

3.3.4 Formung durch Optimierung der p-Norm

Die Optimierung eines auf der Maximumsnorm basierenden Kriteriums, wie in Abschnitt 3.3.3.2 beschrieben, führt zu einer sehr gleichmäßigen Formung des ungewünschten Teils der globalen Impulsantwort. Das in Abschnitt 3.3.3.2 vorgestellte Verfahren von Mei et al. aus [115] benötigt sehr viele Iterationen des Gradientenabstiegsverfahrens zur Optimierung der Zielfunktion.

Von Mertins et al. wurde in [120] eine auf der p-Norm basierende Zielfunktion vorgeschlagen. Die Zielfunktion aus [120] lautet

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad : \quad f(\boldsymbol{h}) \,, \quad f(\boldsymbol{h}) = \log\left(\frac{f_u(\boldsymbol{h})}{f_d(\boldsymbol{h})}\right) \tag{3.57}$$

mit

$$f_d(\boldsymbol{h}) = \|\boldsymbol{g}_d\|_{p_d} = \|\mathrm{diag}\{\boldsymbol{w}_d\}\,\boldsymbol{C}\boldsymbol{h}\|_{p_d} \tag{3.58}$$

und

$$f_u(\boldsymbol{h}) = \|\boldsymbol{g}_u\|_{p_u} = \|\mathrm{diag}\{\boldsymbol{w}_u\}\,\boldsymbol{C}\boldsymbol{h}\|_{p_u} \,, \tag{3.59}$$

wobei $\boldsymbol{C}$ die Faltungsmatrix der Raumimpulsantwort ist und die Vektoren $\boldsymbol{w}_d$ und $\boldsymbol{w}_u$ die Gewichtsfenster $w_d(n)$ und $w_u(n)$ für den gewünschten und den ungewünschten Teil der globalen Impulsantwort enthalten. Wird $p_d = p_u = 2$ gewählt, so ist (3.57) äquivalent zu (3.39); wird $p_d = p_u = \infty$ gewählt, so ist (3.57) äquivalent zu (3.50).

Die Motivation für die Optimierung der p-Norm mit $2 < p < \infty$ ist eine gesteigerte Konvergenzgeschwindigkeit im Vergleich zur Optimierung der Maximumsnorm bei einer gleichmäßigeren Formung als bei der Optimierung einer quadratischen Norm [120]. Üblicherweise werden $10 \leq p_d, p_u \leq 20$ gewählt [75, 81, 120].

Iterative Optimierung der Zielfunktion Die Optimierung von (3.57) erfolgt iterativ durch Anwendung eines Gradientenabstiegsverfahrens [120]. Die Lernregel lautet

$$\boldsymbol{h}^{l+1} = \boldsymbol{h}^{l} - \mu \underbrace{\left(\frac{1}{f_u\big(\boldsymbol{h}^l\big)}\nabla_{\boldsymbol{h}} f_u\big(\boldsymbol{h}^l\big) - \frac{1}{f_d\big(\boldsymbol{h}^l\big)}\nabla_{\boldsymbol{h}} f_d\big(\boldsymbol{h}^l\big)\right)}_{\nabla_{\boldsymbol{h}} f\big(\boldsymbol{h}^l\big)}, \tag{3.60}$$

wobei μ die feste Schrittweite des Gradientenabstiegsverfahrens bezeichnet; l bezeichnet die aktuelle Iteration des Abstiegsverfahrens [120]. Mit

$$\boldsymbol{b}_d = \operatorname{diag}\{\operatorname{sign}\{\boldsymbol{g}_d\}\} \operatorname{diag}\{\boldsymbol{w}_d\} \, |\boldsymbol{g}_d|^{p_d - 1} \,, \tag{3.61}$$

$$\boldsymbol{b}_u = \operatorname{diag}\{\operatorname{sign}\{\boldsymbol{g}_u\}\} \operatorname{diag}\{\boldsymbol{w}_u\} \, |\boldsymbol{g}_u|^{p_u - 1} \,, \tag{3.62}$$

sowie

$$\phi_{f_d} = \sum_{n=0}^{L_g - 1} |g_d(n)|^{p_d} \,, \tag{3.63}$$

$$\phi_{f_u} = \sum_{n=0}^{L_g - 1} |g_u(n)|^{p_u} \tag{3.64}$$

$$\tag{3.65}$$

kann der Gradient der Zielfunktion $f(\boldsymbol{h})$ bezüglich des Zielvektors $\boldsymbol{h}$ als

$$\nabla_{\boldsymbol{h}} f(\boldsymbol{h}) = \frac{1}{\phi_{f_u}} \boldsymbol{C}^T \boldsymbol{b}_u - \frac{1}{\phi_{f_d}} \boldsymbol{C}^T \boldsymbol{b}_d \tag{3.66}$$

berechnet werden.

3.3.4.1 Gewichtsfenster

In [120] wurde vorgeschlagen, die Gewichtsfenster für die Formung der globalen Impulsantwort auf Grundlage von psychoakustischen Eigenschaften des menschlichen Hörsystems zu wählen. Hierzu wird der in Abschnitt 2.2 vorgestellte Effekt der *temporalen Maskierung* ausgenutzt. Ein Schallereignis *maskiert* zeitlich spätere Schallereignisse. Dies bedeutet, dass auf ein Schallereignis folgende Schallereignisse für einen menschlichen Zuhörer nur wahrnehmbar sind, wenn ihre Energie über einer bestimmten Schwelle liegt.

Der exakte Verlauf der Maskierungsschwelle ist abhängig vom Anregungssignal [24]. In [35] wurde eine *durchschnittliche* temporale Maskierungsschwelle definiert; diese ist in Abbildung 2.13 dargestellt.

Auf akustische Übertragungskanäle übertragen bedeutet dies, dass durch die Direktschallkomponente der Impulsantwort eine Maskierungsschwelle induziert wird [120]; für einen menschlichen Zuhörer ist der Nachhall nicht wahrnehmbar, sofern sich dieser unterhalb der Maskierungsschwelle befindet (siehe Abbildung 2.13).

Entsprechend den Überlegungen aus Abschnitt 3.3.3.2 bezüglich der Auswirkung der Gewichtsfenster $w_d(n)$ und $w_u(n)$ auf die Form der globalen Impulsantwort schlagen Mertins et al. in [120] folgende Gewichtsfenster für die Formung vor:

$$\boldsymbol{w}_u = [\underbrace{0, \ldots, 0}_{N_1+N_2}, \underbrace{\boldsymbol{w}_0^T}_{N_3}]^T, \tag{3.67}$$

$$\boldsymbol{w}_d = [\underbrace{0, \ldots, 0}_{N_1} \underbrace{1, \ldots, 1}_{N_2} \underbrace{0, \ldots, 0}_{N_3}]^T, \tag{3.68}$$

mit $N_1 = t_0 f_s$, $N_2 = 0.004\,\mathrm{s}\, f_s$ und $N_3 = L_g - N_1 - N_2$, wobei f_s die Abtastrate und t_0 die Zeit des Eintreffens der Direktschallkomponente bezeichnen. Das Fenster $\boldsymbol{w}_0$ ist als

$$w_0(n) = 10^{\frac{3}{\log(N_0/(N_1+N_2))} \log\left(\frac{n}{N_1+N_2}\right)+0.5} \tag{3.69}$$

mit $N_0 = (0.2\,\mathrm{s} + t_0) f_s$ definiert; der Zeitindex n läuft in (3.69) von $N_1 + N_2 + 1$ bis $L_g - 1$ [120]. Der Kehrwert von (3.69) entspricht der durschnittlichen temporalen Maskierungsschwelle nach [35]. Eine grafische Darstellung der Gewichtsfenster ist in Abbildung 3.15 gegeben.

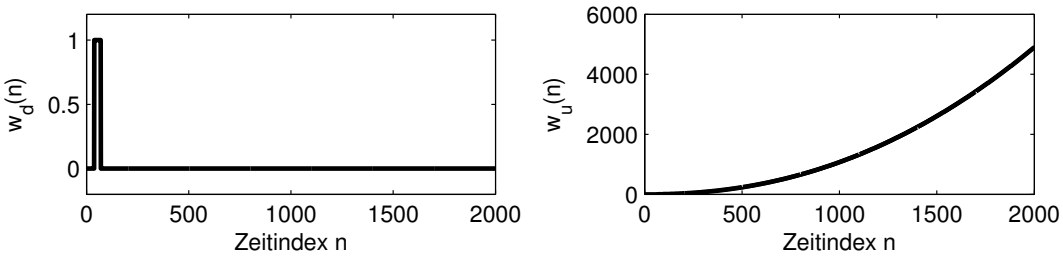

Abbildung 3.15: Darstellung der von Mertins et al. in [120] empfohlenen Gewichtsfenster. Die Gewichtsfenster sind entsprechend der von Fielder in [34, 35] formulierten durchschnittlichen temporalen Maskierungsschwelle des menschlichen auditorischen Systems gewählt.

Beispiele Zur Demonstration der Impulsantwortformung durch die Optimierung der p-Norm nach (3.57) wird die in Abbildung 3.3 dargestellte simulierte Raumimpulsantwort verwendet. Für die Beispiele in diesem Abschnitt wurde jeweils ein Filter $h(n)$ der Länge $L_h = 1000$ entworfen. Die Schrittweite des Gradientenabstiegsverfahrens wurde auf $\mu = 10^{-5}$ gesetzt und es wurden $2 \cdot 10^5$ Iterationen des Gradientenabstiegsverfahrens durchgeführt. Des Weiteren wurden $p_d = 20$ und $p_u = 10$ als Parameter für die p-Norm in (3.57) gewählt.

Für das erste Beispiel werden die Rechteckfenster nach (3.41) und (3.42) (siehe Abbildung 3.9) verwendet. Das Ergebnis der Optimierung ist in Abbildung 3.16 dargestellt.

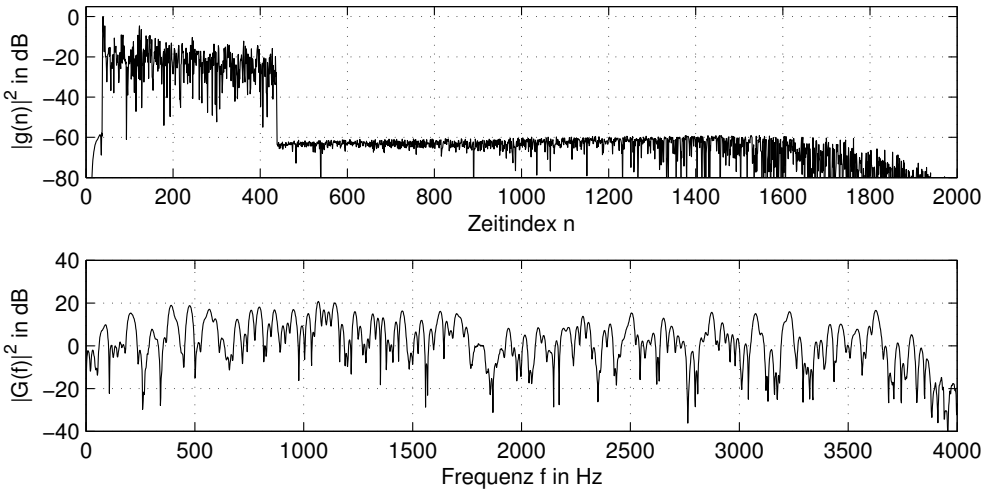

Abbildung 3.16: Darstellung der nach dem Verfahren aus [120] entzerrten Raumimpulsantwort im Zeit- (oben) und Frequenzbereich (unten). Die Gewichtsfenster wurden, wie in [85] vorgeschlagen, als Rechteckfenster gewählt.

Für das zweite Beispiel wurden die Gewichtsfenster wie in [120] vorgeschlagen nach der temporalen Maskierungsschwelle des menschlichen Hörsystems [35] gewählt (siehe Abbildung 3.15). Das Ergebnis der Optimierung von (3.57) mit diesen Fenstern ist in Abbildung 3.17 dargestellt; die gestrichelte schwarze Linie notiert die temporale Maskierungsschwelle nach [35].

Zum Vergleich des auf der p-Norm basierenden Verfahrens mit einer quadratischen Norm wurde das gleiche Experiment mit $p_d = p_u = 2$ durchgeführt; die Gewichtsfenster wurden entsprechend der temporalen Maskierungsschwelle gewählt. Das Ergebnis der Optimierung von (3.57) beziehungsweise (3.39) ist in Abbildung 3.18 dargestellt.

3.3.5 Zusammenfassung

In diesem Abschnitt wurden Methoden zur Verkürzung und Formung von Raumimpulsantworten ohne vorgegebenes Zielsystem vorgestellt. Das in Abschnitt 3.3.4 präsentierte Verfahren zur Formung durch die Optimierung einer auf der p-Norm basierenden Zielfunktion bildet die Grundlage der im Rahmen dieser Arbeit entwickelten Methoden.

Im Vergleich zu der quadratischen Zielfunktion aus Abschnitt 3.3.1 zeichnen sich die auf der Maximums- beziehungsweise auf der p-Norm basierenden Algorithmen

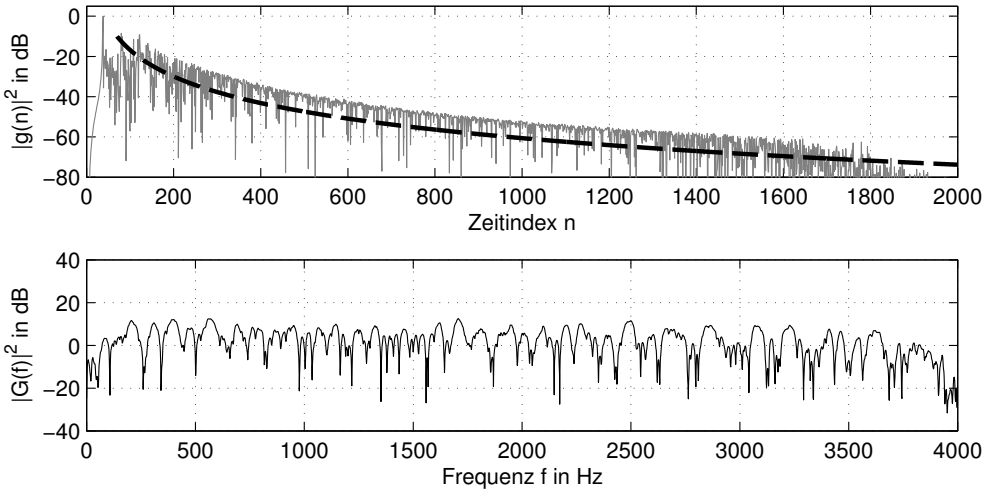

Abbildung 3.17: Darstellung der nach dem Verfahren aus [120] entzerrten Raumimpulsantwort im Zeit- (oben) und Frequenzbereich (unten). Die Gewichtsfenster wurden wie in [120] vorgeschlagen im Einklang mit der temporalen Maskierungsschwelle des menschlichen Hörsystems [35] gewählt. Die gestrichelte schwarze Linie markiert die durchschnittliche temporale Maskierungsschwelle [35].

dadurch aus, dass die Formung im Zeitbereich üblicherweise zu keinen starken spektralen Verzerrungen führt.

Beim Vergleich der Ergebnisse bei Optimierung der p-Norm (Abbildung 3.16) und bei Optimierung der Maximumsnorm (Abbildung 3.14) ist zu erkennen, dass die Formung durch den auf der p-Norm basierenden Algorithmus bei einer um den Faktor 100 geringeren Anzahl an Iterationen vergleichbare Ergebnisse liefert. Die gesteigerte Konvergenzgeschwindigkeit wurde in [120] als Eigenschaft des auf der p-Norm basierenden Algorithmus aufgeführt.

Ein Vergleich von Abbildung 3.17 und Abbildung 3.18 zeigt, dass die Optimierung der p-Norm mit $p > 2$ zu einer gleichmäßigeren Formung als die Optimierung einer quadratischen Norm führt; darüber hinaus weist der Betragsfrequenzgang keine extremen Verzerrungen auf.

Nicht weiter betrachtet wird im Rahmen diese Arbeit das mehrkanalige Verfahren von Kodrasi et al. [93]. Hierbei wird in einem ersten Schritt eine Kanalverkürzung nach dem mehrkanaligen MSSNR Verfahren (3.39) durchgeführt [84]. In einem zweiten Schritt wird ein vorgegebenes Zielsystem approximiert. Hierbei wird der Entzerrer in dem von den Eigenvektoren des Rayleigh-Quotienten (3.39) aufgespannten Unterraums gesucht.

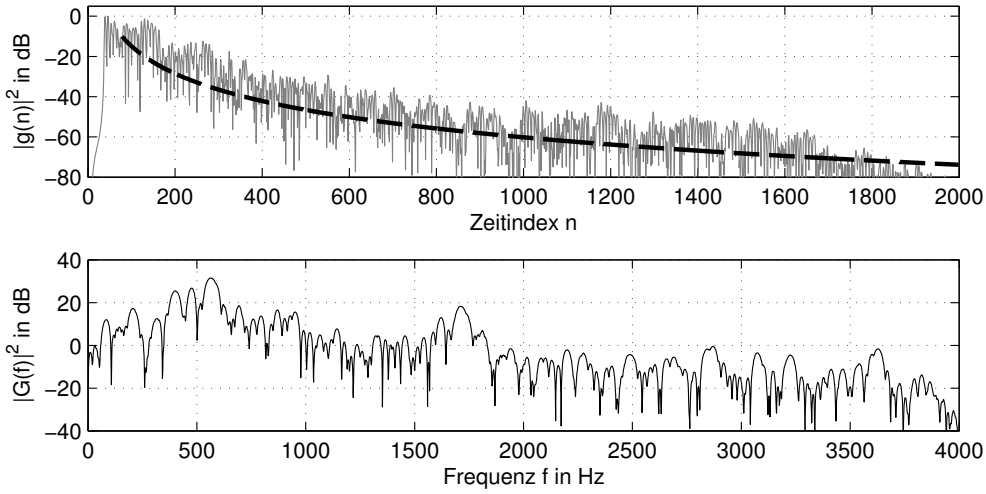

Abbildung 3.18: Darstellung der entzerrten Raumimpulsantwort im Zeit- (oben) und Frequenzbereich (unten) durch Optimierung einer quadratischen Norm. Durch die Wahl der Gewichtsfenster wurde, wie in [120] vorgeschlagen, gefordert, dass das zeitliche Abklingverhalten der temporalen Maskierungsschwelle (gestrichelte schwarze Linie) nach [35] folgt.

3.4 Robuste Kanalkompensation

Für die kanalbasierte Raumhallkompensation wird ein Entzerrer auf Grundlage einer Raumimpulsantwort entworfen. Von großer Relevanz für die praktische Nutzbarmachung der Verfahren ist die Robustheit gegenüber einer fehlerbehafteten Schätzung der Raumimpulsantwort sowie gegenüber räumlicher Fehlplatzierung. Radlović et al. empfehlen, von eine Entzerrung abzusehen, wenn nicht garantiert werden kann, dass sich der Zuhörer an der Referenzposition befindet [139, 140].

Das generelle Vorgehen zum Entwurf robuster Entzerrer ist, den akustischen Kanal als potentiell fehlerbehaftet zu modellieren. Üblicherweise wird zur Modellierung der Systemabweichungen ein additiver Störterm verwendet; die übliche Notation zur Beschreibung von perturbierten Raumimpulsantworten wird in Abschnitt 3.4.1 eingeführt. Für eine robuste Entzerrung wird der Perturbationsterm in den Filterentwurf integriert.

Eine weitere Möglichkeit zur Erhöhung der räumlichen Robustheit ist der von Mei et al. in [114] verfolgte Ansatz. Hierbei wird ein Quellensignal von mehreren Lautsprechern gleichzeitig wiedergegeben. Es werden mehrere Raumimpulsantworten innerhalb eines Volumens gemessen und die Vorfilter werden derart entworfen, dass alle Raumimpulsantworten gleichmäßig geformt werden. Wurde bei der Messung der Raumimpulsantworten das räumliche Abtasttheorem für Raumimpulsantworten [1] eingehalten, so kann eine

gleichmäßige Entzerrung innerhalb des Volumens erreicht werden [114]. Es hat sich hierbei gezeigt, dass die alleinige Verwendung mehrere Lautsprecher zwar den Grad der Entzerrung, nicht jedoch den räumlichen Gültigkeitsbereich verbessert [131]. In Abschnitt 3.4.5 wird der Ansatz nach Mei et al. [114] vorgestellt.

Weitere Ansätze zum Entwurf räumlich robuster Entzerrer werden in Abschnitt 3.4.6 vorgestellt; in Abschnitt 3.4.2 wird explizit auf die Robustheit gegen Kanalfehlschätzungen eingegangen.

3.4.1 Beschreibung von fehlerbehafteten Kanälen

Die übliche Vorgehensweise zum Entwurf robuster Entzerrer ist, die zu entzerrende Raumimpulsantwort als potentiell fehlerbehaftet zu modellieren [61, 90, 94, 169]. In Abbildung 3.19 ist schematisch dargestellt, wie eine additive Perturbation $p(n)$ parallel zu der gemessenen Raumimpulsantwort $c(n)$ geschaltet ist, um die zu entzerrende Raumimpulsantwort $\hat{c}(n)$ zu beschreiben.

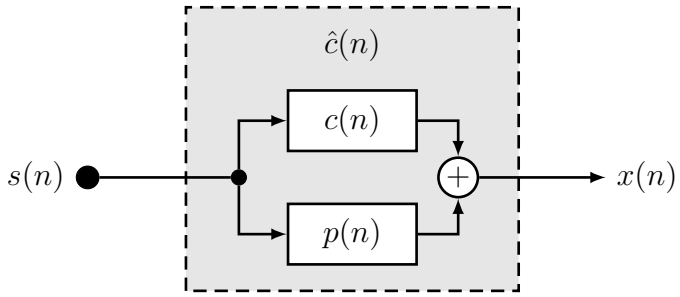

Abbildung 3.19: Schematische Darstellung eines fehlerbehafteten Kanals $\hat{c}(n) = c(n) + p(n)$. Die Perturbation $p(n)$ kann in den Filterentwurf integriert werden, um beispielsweise die Robustheit der Entzerrer gegenüber Kanalfehlschätzungen zu erhöhen.

Die zu entzerrende Raumimpulsantwort wird wie in Abbildung 3.19 dargestellt als

$$\hat{c}(n) = c(n) + p(n) \tag{3.70}$$

beschrieben [94]. Je nach Anwendungsfall kann angenommen werden, dass die Raumimpulsantwort *vollständig* bekannt ist

$$p(n) = 0 \Longleftrightarrow \hat{c}(n) = c(n)\,, \tag{3.71}$$

oder dass ein Fehler vorliegt, also $p(n) \neq 0$, der durch eine fehlerbehaftete Messung oder räumliche Fehlplatzierung verursacht wurde.

Der Störterm $p(n)$ wird üblicherweise über statistische Eigenschaften modelliert. Ein Überblick über die wichtigsten Arten der Modellierung wird in den Abschnitten 3.4.2 und 3.4.6 gegeben. Im Folgenden wird der Entwurf robuster Entzerrer durch die Approximation eines gewünschten Zielsystems beschrieben. Der Signalflussgraph zu dieser Problemstellung ist in Abbildung 3.20 dargestellt.

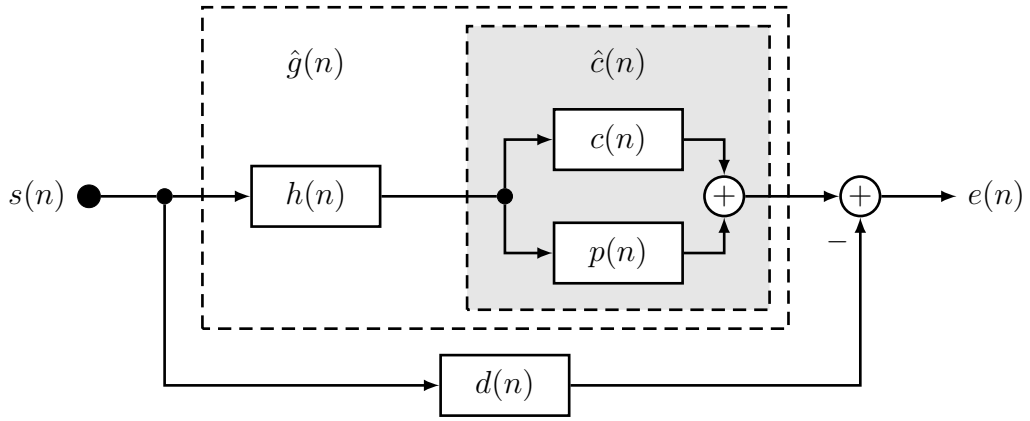

Abbildung 3.20: Signalflussgraph für die Kompensation eines fehlerbehafteten Kanals $\hat{c}(n) = c(n) + p(n)$ durch ein Filter $h(n)$ mit vorgegebenem Zielsystem $d(n)$. Der Entzerrer $h(n)$ wird durch die Minimierung der Fehlerenergie $\mathrm{E}\{e^2(n)\}$ entworfen.

3.4.2 Robustheit gegen Kanalfehlschätzungen

Zum Entwurf von Filtern, die robust gegenüber Kanalfehlschätzungen sind, wird das Ausgabesignal $\hat{y}(n)$ des Gesamtsystems betrachtet:

$$\begin{aligned}\hat{y}(n) &= s(n) * (\hat{c}(n) * h(n)) \\ &= s(n) * ((c(n) + p(n)) * h(n)) \\ &= \underbrace{s(n) * c(n) * h(n)}_{y(n)} + \underbrace{s(n) * p(n) * h(n)}_{\tilde{y}(n)},\end{aligned} \tag{3.72}$$

wobei der perturbierte Kanal $\hat{c}(n)$ wie in (3.70) modelliert wird; $s(n)$ ist das Quellensignal und $h(n)$ bezeichnet den zu entwerfenden Entzerrer. Der erste Term in (3.72) entspricht dem gewünschten Signal $y(n)$, welches korrekt entzerrt wurde. Der zweite

Term in (3.72) entspricht dem ungewünschten Signalanteil $\tilde{y}(n)$, welcher durch die fehlerbehaftete Kanalschätzung entstanden ist [94].

Das in der Literatur übliche Vorgehen zum Entwurf robuster Entzerrer ist, die Energie des Störsignals

$$\tilde{y}(n) = s(n) * p(n) * h(n) \tag{3.73}$$

zu minimieren [61, 94]. Um die Energie des Störsignals zu minimieren, kann beispielsweise gefordert werden, dass der Entzerrer minimale Energie hat und somit der Übertragungskanal $p(n) * h(n)$ minimale Energie aufweist.

Robuste Entzerrer durch Minimierung der Filterenergie Zur Minimierung der Energie des Entzerrers wird das Optimierungsproblem (3.18) durch einen Regularisierungsterm erweitert:

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad : \quad \|\boldsymbol{C}\boldsymbol{h} - \boldsymbol{d}\|_2^2 + \alpha \|\boldsymbol{h}\|_2^2 , \tag{3.74}$$

wobei $\alpha \geq 0$ der Gewichtsfaktor für den Regularisierungsterm ist und $\boldsymbol{C}$ die Faltungsmatrix der gemessenen Raumimpulsantwort $c(n)$; $\boldsymbol{d}$ ist das zu approximierende Zielsystem. Die Lösung zu (3.74) ist durch

$$\boldsymbol{h} = \left(\boldsymbol{C}^T\boldsymbol{C} + \alpha\boldsymbol{I}\right)^{-1}\boldsymbol{C}^T\boldsymbol{d} \tag{3.75}$$

in geschlossener Form gegeben, wobei $\boldsymbol{I}$ die Einheitsmatrix bezeichnet [61–63, 88].

Modellierung der Perturbationen Im Allgemeinen ist es nicht möglich, die durch die Kanalidentifikation gemachten Fehler statistisch exakt zu beschreiben [95]. In [169] wurde systematisch von Zhang et al. untersucht, wie sich die Robustheit der Entzerrer gegenüber dem bei der blinden Kanalidentifikation gemachten Fehler erhöhen lässt. Hierzu simulierten Zhang et al. nach der Spiegelmethode [2] ein sechskanaliges akustisches Übertragungssystem und verwendeten den Algorithmus aus [65] zur Schätzung der einzelnen Kanäle. Ein Vergleich der simulierten Raumimpulsantworten mit den geschätzten Impulsantworten zeigte, dass sich der Fehler $p(n)$ (siehe Gleichung (3.70)) im Falle der blinden Kanalidentifizierung als Zufallsfolge mit einem exponentiellen Abfall, der dem der Raumimpulsantwort gleicht, beschreiben lässt [169].

Zhang et al. modellieren den Systemfehler als

$$p(n) = \alpha b(n)\, e^{-\Delta n}, \tag{3.76}$$

wobei $b(n)$ eine mittelwertfreie, unkorrelierte gaussförmige Rauschsequenz mit Varianz $\sigma^2_{b(n)} = 1$ ist, und Δ ist über

$$\Delta = \frac{3\ln(10)}{T_{60} f_s} \tag{3.77}$$

durch die Nachhallzeit T_{60} und die Abtastrate f_s bestimmt; der Faktor α bestimmt die Fehlerenergie und kann als Gewichtsfaktor für die Regularisierung verstanden werden.

Zhang et al. schlagen das folgende Optimierungsproblem als direkte Erweiterung von (3.23) vor, um robuste Entzerrer zu entwerfen:

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad : \quad \mathrm{E}\big\{ \|\boldsymbol{W}((\boldsymbol{C}+\boldsymbol{P})\,\boldsymbol{h} - \boldsymbol{d})\|_2^2 \big\}\,, \tag{3.78}$$

wobei $\boldsymbol{C}$ die Faltungsmatrix der fehlerbehafteten Schätzung der Raumimpulsantwort ist und $\boldsymbol{P}$ die Faltungsmatrix der Perturbation $p(n)$ ist.

Die Lösung zu (3.78) ist durch

$$\boldsymbol{h} = \big(\boldsymbol{C}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{C} + \mathrm{E}\big\{\boldsymbol{P}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{P}\big\}\big)^{-1}\widehat{\boldsymbol{C}}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{d} \tag{3.79}$$

gegeben. Die Matrix $\boldsymbol{R} = \mathrm{E}\big\{\boldsymbol{P}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{P}\big\}$ lässt sich unter den von Zhang et al. in [169] gemachten Annahmen (3.76) als Diagonalmatrix mit den Elementen

$$r(j) = \alpha^2 \sum_{i=0}^{L_c-1} w^2(i+j-1)\, e^{-2\Delta i}, \quad j = 1, \ldots, L_h, \tag{3.80}$$

auf der Hauptdiagonalen berechnen. Je nachdem, wie die Perturbation $p(n)$ modelliert wird, ergeben sich für die Matrix $\boldsymbol{R}$ andere Ausdrücke. Wird $p(n)$ als mittelwertfreies stationäres weißes Rauschen modelliert und $\boldsymbol{W}$ als Einheitsmatrix gewählt, so ergibt sich beispielsweise die in (3.75) gegebene Lösung für (3.79) [61].

3.4.3 Räumliche Robustheit

Eng mit dem in Abschnitt 3.4.2 erläuterten Problem der fehlerbehafteten Kanalidentifikation verwandt ist die Robustheit der Entzerrung im Falle von räumlicher Fehlplatzierung. Wird ein Entzerrer für eine Referenzposition entworfen und auf eine Raumimpulsantwort abseits der Referenzposition angewendet, so ist üblicherweise mit einer Verschlechterung der akustischen Übertragungseigenschaften des Gesamtsystems zu rechnen [128, 139]. Es wird sogar empfohlen, gar keine Entzerrung vorzunehmen, wenn nicht garantiert werden kann, dass kein räumlicher Versatz vorliegt [139, 140].

Für die praktische Nutzbarmachung der kanalbasierten Raumhallkompensation ist die räumliche Robustheit der Entzerrer von essentieller Bedeutung, da es nicht möglich ist, einen menschlichen Zuhörer exakt an einer Raumposition zu fixieren.

In Abschnitt 3.4.4 werden die Untersuchungen von Radlović et al. zur Auswirkung von räumlicher Fehlplatzierung vorgestellt. In den Abschnitten 3.4.5 und 3.4.6 wird ein Überblick über mehrere Ansätze zur Vergrößerung des räumlichen Gültigkeitsbereiches der Entzerrer gegeben.

Für die Betrachtung der räumlichen Robustheit wird in der Literatur üblicherweise lediglich von einer Entzerrung des Betragsfrequenzgangs ausgegangen. Im Rahmen dieser Arbeit wurden Verfahren erarbeitet, die auch im Falle von räumlicher Fehlplatzierung sowohl einen flachen Betragsfrequenzgang als auch eine Enthallung (durch die Formung im Zeitbereich) ermöglichen.

3.4.4 Untersuchungen zur räumlichen Robustheit

Die räumliche Robustheit von Entzerrern wurde systematisch von Radlović et al. untersucht [139, 140]. In dem von Radlović et al. betrachteten Szenario wird von einer *perfekten* Entzerrung (3.1) der Übertragungsfunktion an der Referenzposition ausgegangen. Nach dem Filterentwurf wird das Mikrofon aus der Referenzposition bewegt und der entworfene Entzerrer wird auf die geänderte Übertragungsfunktion angewendet.

Für den Entwurf des Fehlermodels betrachten Radlović et al. die statistischen Eigenschaften von akustischen Übertragungsfunktionen in geschlossenen Räumen [145]. Sind gewisse Bedingungen erfüllt, so kann das Schallfeld innerhalb eines Raumvolumens als Zufallsfunktion beschrieben werden [139]. Die hierfür nötigen Bedingungen schränken die praktische Bedeutung des Szenarios in keinster Weise ein:

1. Die Abmessungen des Raumes müssen größer als die betrachteten Wellenlängen sein.

2. Die betrachteten Frequenzen liegen über der *Schröder-Frequenz für große Räume* [148], die durch
$$f_S = 2000\sqrt{\frac{T_{60}}{V}}\,\mathrm{Hz} \tag{3.81}$$
gegeben ist. Dabei ist T_{60} die Nachhallzeit des Raumes [98] und V das Volumen des Raumes in Kubikmetern (m^3). Für einen *normalen* Raum mit einem Volumen von $60\,\mathrm{m}^3$ und einer Nachhallzeit von $T_{60} = 0.6\,\mathrm{s}$ beträgt die Schröder-Frequenz

$f_S = 200\,\mathrm{Hz}$; oberhalb dieser Frequenz können die Übertragungsfunktionen als statistische Prozesse angesehen werden [139, 148].

3. Sowohl die Schallquelle als auch der Empfänger befinden sich innerhalb des Raumes und haben einen ausreichend großen Abstand zu den Wänden. Der nötige Abstand beträgt mindestens eine halbe Wellenlänge der niedrigsten betrachteten Frequenz. Wird die Schröder-Frequenz als niedrigste Frequenz angenommen, so beträgt der nötige Abstand bei $f_S = 200\,\mathrm{Hz}$ etwa $0.86\,\mathrm{m}$ (Wellenlänge $\lambda = v/f_S$ mit der Schallgeschwindigkeit $v = 344\,\mathrm{m/s}$).

Sind diese Bedingungen erfüllt, so lassen sich die akustischen Übertragungsfunktionen als stochastische Prozesse beschreiben, deren Eigenschaften durch das Raumvolumen, die Nachhallzeit des Raumes und den Schalldruck bestimmt werden [103, 139, 167]. Die gleichen statistischen Annahmen gelten entsprechend auch für das Inverse der Übertragungsfunktion [139].

Als weitere Bedingung soll gelten, dass sich das Mikrofon im Diffusfeld befindet, die Energie der Reflexionen im Vergleich zur Direktschallkomponente also nicht zu vernachlässigen sind [139]. Die hierfür notwendige minimale Entfernung R_{min} in Metern zwischen Schallquelle und Mikrofon ist dabei über die Schröder-Frequenz (3.81) durch

$$R_{\text{min}} = \frac{200}{f_S} \quad [\mathrm{m}] \tag{3.82}$$

gegeben.

Frequenzabhängiges Fehlermaß Bezeichne $C(f)$ die Übertragungsfunktion von einem Lautsprecher zu einer Referenzposition, wobei f die Frequenz in Hertz ist. $H(f)$ bezeichne den *perfekten* Entzerrer für die Übertragungsfunktion zur Referenzposition und ist durch

$$H(f) = \frac{1}{C(f)} \tag{3.83}$$

gegeben. Des Weiteren bezeichne $\widehat{C}(f)$ die Übertragungsfunktion zum Ort der Entzerrung. Der mittlere quadratische Fehler, der sich ergibt, wenn der Entzerrer $H(f)$ auf $\widehat{C}(f)$ angewendet wird, ist durch

$$U(f) = \mathrm{E}\left\{ \left| \widehat{C}(f)\, H(f) - 1 \right|^2 \right\} \tag{3.84}$$

gegeben. Mit der Beschreibung der Veränderung der Raumimpulsantwort als additive Perturbation (3.70) kann (3.84) als

$$U(f) = \mathrm{E}\left\{\left|\left(C(f) + P(f)\right) H(f) - 1\right|^2\right\} \tag{3.85}$$

geschrieben werden; $P(f)$ bezeichnet hierbei die Fourier-Transformierte des Perturbationsterms $p(t)$. Im Falle einer Entzerrung an der Referenzposition gilt $P(f) = 0$ und damit $U(f) = 0$ für alle Frequenzen f [139].

Frequenzabhängige Fehlerenergie bei Fehlplatzierung Wird das Mikrofon um die Entfernung D aus der Referenzposition bewegt, so ist das Verhalten des mittleren quadratischen Fehlers aus (3.84) durch

$$U(f) \cong 2 - 2\,\underbrace{\frac{\sin(\kappa D)}{\kappa D}}_{\mathrm{si}(\kappa D)} \tag{3.86}$$

gegeben, wobei die Wellenzahl κ durch $\kappa = 2\pi/\lambda = 2\pi f/v$ gegeben ist [139]; λ bezeichnet hierbei die Wellenlänge und v die Schallgeschwindigkeit.

Gleichung (3.86) besagt, dass der größte Fehler bei hohen Frequenzen und/oder einer großen Fehlplatzierung zu erwarten ist, da $\mathrm{si}(\kappa D)$ für große Werte von κD klein wird. Des Weiteren ist die frequenzabhängige Fehlerenergie unabhängig von der Größe des Raumes – sofern die in diesem Abschnitt beschriebenen Bedingungen erfüllt sind [139].

Exemplarische Ergebnisse Als Ergebnisse präsentieren Radlović et al. in ihren Arbeiten die Gegenüberstellung ihres theoretischen Modells für die frequenzabhängige Fehlerenergie (3.86) und die sich aus einer Simulation ergebenen Werte [139, 140]. Hierfür werden Raumimpulsantworten in einem Raum der Größe $[6.4 \times 5 \times 4]\ \mathrm{m}^3$ nach der Spiegelbildmethode [2] simuliert. Die Nachhallzeit des Raumes beträgt $T_{60} = 0.45\,\mathrm{s}$ und die Distanz zwischen Lautsprecher und der Referenzposition beträgt $R = 3\,\mathrm{m}$. Sowohl der Lautsprecher als auch alle betrachteten Mikrofonpositionen sind jeweils weiter als einen Meter von den Wänden des Raumes entfernt.

Die von Radlović et al. präsentierten Ergebnisse zeigen die Entwicklung der Fehlerenergie für $f = 2000\,\mathrm{Hz}$ in Abhängigkeit der Entfernung zur Referenzposition in Vielfachen der Wellenlänge $\lambda_{2000\,\mathrm{Hz}} = 0.17\,\mathrm{m}$.

Es wird zunächst die Übertragungsfunktion $C(f)$ zu der Referenzposition simuliert und hieraus der *perfekte* Entzerrer $H(f) = \frac{1}{C(f)}$ berechnet. Danach werden in unter-

schiedlichen Abständen von der Referenzposition zufällig verteilt jeweils $N = 200$ weitere Übertragungsfunktionen ermittelt. Für $f = 2000\,\mathrm{Hz}$ wird jetzt die durchschnittliche Fehlerenergie nach (3.84) berechnet. Die Ergebnisse sind in Abbildung 3.21 dargestellt. Hierbei ist in rot (gestrichelt) die nach dem Modell (3.86) von Radlović et al. geschätzte Fehlerenergie für $f = 2000\,\mathrm{Hz}$ in Abhängigkeit der Fehlplatzierung aufgetragen. Die blaue Linie in Abbildung 3.21 stellt die nach (3.84) auf Grundlage der simulierten Übertragungsfunktionen ermittelte Fehlerenergie dar.

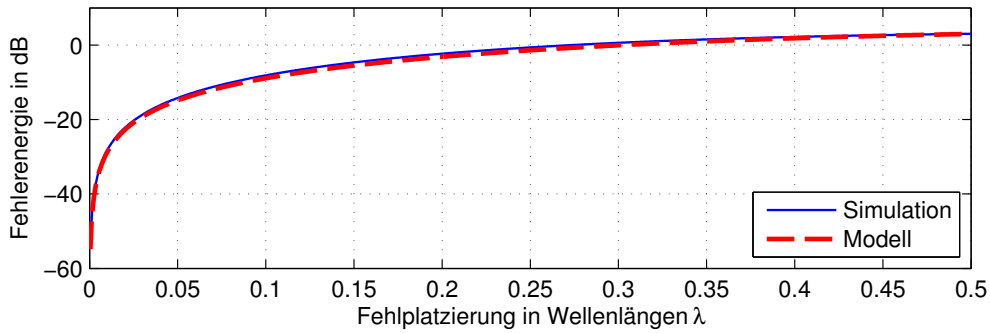

Abbildung 3.21: Gegenüberstellung der aus Simulationen ermittelten Fehlerenergie in Abhängigkeit der Fehlplatzierung für $f = 2000\,\mathrm{Hz}$ (blau) und dem erwarteten Verlauf der Fehlerenergie (rot, gestrichelt) bei einer perfekten Entzerrung am Referenzpunkt [139, 140].

Die in [12] von Bharitkar et al. durchgeführten Untersuchungen bestätigen das von Radlović et al. in [139] erarbeitete Modell zur statistischen Beschreibung der frequenz- und entfernungsabhängigen Fehlerenergie.

3.4.5 Erhöhung der Robustheit gegenüber Fehlplatzierung durch multiple Messungen

Neben der Robustheit gegen Kanalfehlschätzungen gibt es in der Literatur mehrere Ansätze zur Erhöhung der Robustheit gegenüber räumlicher Fehlplatzierung. Die Veränderung des akustischen Kanals durch eine Fehlplatzierung kann generell als additive Störung des Kanals aufgefasst werden; wird der additive Störterm in erster Näherung als unkorreliertes gaussförmiges Rauschen angenommen, ergeben sich die gleichen Terme wie in Abschnitt 3.4.2.

In [127] wurde vorgeschlagen, die Entzerrer getrennt für mehrere Raumvolumina zu entwerfen und entsprechend der Position des Zuhörers anzuwenden.

In [114] wurde eine räumlich robuste Nachhallreduktion durch die Abtastung eines Volumens mit Mikrofonen erreicht. Die hierfür entwickelte Theorie basiert auf der Abtastung der plenakustischen Funktion unter Beachtung des räumlichen Abtasttheorems [1]. Das im Rahmen dieser Arbeit entwickelte Formelwerk erlaubt den Entwurf räumlich robuster Entzerrer nach dem von Mei et al. in [114] vorgeschlagenen Mehrpositions-Ansatz; diese Methode soll an dieser Stelle vorgestellt werden.

Räumliches Abtasttheorem für Raumimpulsantworten Die zeitkontinuierliche Raumimpulsantwort von einem Sender am Ort $\boldsymbol{x}' = [x', y', z']^T$ zu einem Empfänger am Punkt $\boldsymbol{x} = [x, y, z]^T$ wird als $c_{\boldsymbol{x}'\boldsymbol{x}}(t)$ bezeichnet. Unter Beachtung des Abtasttheorems [149] muss die Zeitdarstellung der Raumimpulsantwort mit (mindestens) dem Doppelten der höchsten in den zu übertragenden Signalen vorkommenden Frequenz abgetastet werden; diese Frequenz wird als f_0 bezeichnet.

Bei einer fixen Position des Lautsprechers ist $c_{\boldsymbol{x}'\boldsymbol{x}}(t)$ zu jedem Zeitpunkt t eine örtlich-bandbegrenzte Funktion in der Position $\boldsymbol{x}$ des Zuhörers. Hierbei bestimmt die temporale Abtastrate f_0 über f_0/v die höchste räumliche Frequenz, wobei v die Schallgeschwindigkeit ist [1, 114].

Die räumliche Abtastrate $f_{\text{s.s.}}$ der plenakustischen Funktion (also der einzelnen Raumimpulsantworten) muss

$$f_{\text{s.s.}} \geq 2\frac{f_0}{v} \tag{3.87}$$

erfüllen [1]. In [114] wird allgemein von

$$f_{\text{s.s}} > \frac{f_{\text{t.s.}}}{v} \tag{3.88}$$

ausgegangen, wobei $f_{\text{t.s.}}$ die Abtastrate der Zeitbereichsdarstellung bezeichnet.

Bezeichne $\boldsymbol{i} = [i, j, k]^T$ die (räumlich diskrete) Position des Lautsprechers und $\boldsymbol{l} = [l, m, n]^T$ die (räumlich diskrete) Position des Empfängers, so kann unter Einhaltung der räumlichen Abtastbedingung (3.87) die Raumimpulsantwort an beliebigen Positionen $\boldsymbol{x}$ interpoliert werden [1, 79, 114]:

$$c_{\boldsymbol{i}\boldsymbol{x}}(t) = \sum_{\boldsymbol{l}} \phi(\boldsymbol{x}, \boldsymbol{l}) \cdot c_{\boldsymbol{i}\boldsymbol{l}}(t)\,, \tag{3.89}$$

wobei

$$\phi(\boldsymbol{x}, \boldsymbol{l}) \equiv \phi(x - l\Delta_x, y - m\Delta_y, z - n\Delta_z) \tag{3.90}$$

der Interpolationskern ist. Die Bezeichner Δ_x, Δ_y und Δ_z sind die räumlichen Abtastintervalle in den drei Raumdimensionen. Als Interpolationskern im Raum wird die dreidimensionale $\mathrm{si}(\cdot)$ Funktion (vergleiche *Abtastung zeitkontinuierlicher Signale* in [119]) verwendet [114]:

$$\phi(\boldsymbol{x}) = \mathrm{si}\left(\frac{\pi x}{\Delta_x}\right) \cdot \mathrm{si}\left(\frac{\pi y}{\Delta_y}\right) \cdot \mathrm{si}\left(\frac{\pi z}{\Delta_z}\right), \tag{3.91}$$

wobei $\mathrm{si}(a) = \frac{\sin(a)}{a}$.

Zusammenfassend besagt das räumliche Abtasttheorem für Raumimpulsantworten, dass die räumliche Grenzfrequenz durch die maximale zeitliche Frequenz in den betrachteten Signalen bestimmt wird [1, 114].

Räumlich robuste Formung von Raumimpulsantworten Mit dem räumlichen Abtasttheorem für Raumimpulsantworten [1] und $h_{\boldsymbol{i}}(t)$ als Vorfilter für den Lautsprecher an Position $\boldsymbol{i}$ ist die globale Impulsantwort an der Position $\boldsymbol{x}$ durch

$$\begin{aligned} g_{\boldsymbol{x}}(t) &= \sum_{\boldsymbol{i}} h_{\boldsymbol{i}}(t) * c_{\boldsymbol{i}\boldsymbol{x}}(t) \\ &= \sum_{\boldsymbol{l}} g_{\boldsymbol{l}}(t) \cdot \phi(\boldsymbol{x}, \boldsymbol{l}) \end{aligned} \tag{3.92}$$

gegeben, wobei $g_{\boldsymbol{l}}(t) = \sum_{\boldsymbol{i}} h_{\boldsymbol{i}} * c_{\boldsymbol{i}\boldsymbol{l}}(t)$ die globale Impulsantwort an der Position $\boldsymbol{l}$ innerhalb des Abhörvolumens ist.

Im Folgenden werden die geformten Impulsantworten betrachtet. Werden die räumlichen Charakteristika von Raumimpulsantworten betrachtet [1, 114, 139], so erfüllen die geformten Raumimpulsantworten folgende Hypothesen:

1. Für einen Zeitpunkt t kann $g_{\boldsymbol{l}}(t)$ als stationäres Feld angenommen (beziehungsweise mit vernachlässigbar kleinem Fehler als stationäres Feld approximiert werden). Damit gilt $\mathrm{E}\{g_{\boldsymbol{l}}(t)\} = \mathrm{const}(t)$ und die Korrelationsfunktion

$$r_g(t, \boldsymbol{l}, \boldsymbol{i}) = \mathrm{E}\{g_{\boldsymbol{l}}(t)\, g_{\boldsymbol{i}}(t)\} \tag{3.93}$$

ist lediglich von den Differenzen $l - i$, $m - j$, und $n - k$ (beziehungsweise in Vektor-Notation $\boldsymbol{l} - \boldsymbol{i}$) abhängig. Somit lässt sich (3.93) kurz als

$$\mathrm{E}\{g_{\boldsymbol{l}}(t)\, g_{\boldsymbol{i}}(t)\} = r_g(t, \boldsymbol{l} - \boldsymbol{i}) \tag{3.94}$$

schreiben.

2. Der Betrag $|g_{\boldsymbol{l}}(t)|$ ist nach oben hin begrenzt:

$$|g_{\boldsymbol{l}}(t)| \leq g_0(t)\,, \quad \text{für} \;\; t_1 < t < t_2, \tag{3.95}$$

wobei $g_0(t) > 0$ die obere Grenze bezeichnet (vergleiche (3.53) in Abschnitt 3.3.3.2). Gleichung (3.95) impliziert, dass

$$r_g(t, \mathbf{0}) = \mathrm{E}\big\{g_{\boldsymbol{l}}^2(t)\big\} \leq \mathrm{E}\big\{g_0^2(t)\big\}\,, \quad \text{für} \;\; t_1 < t < t_2, \tag{3.96}$$

wobei das Zeitintervall (t_1, t_2) beispielsweise durch die temporalen Maskierungseigenschaften des menschlichen auditorischen Systems bestimmt wird [35, 114].

Als Zeitintervall (t_1, t_2) wird, wie in [120] vorgeschlagen, der Zeitbereich ab 4 Millisekunden nach der Direktschallkomponente bis $t \to \infty$ betrachtet; also der *ungewünschte* Teil der globalen Impulsantwort (siehe Abschnitt 3.3.4.1). Dieser Abschnitt der globalen Impulsantworten soll entsprechend der temporalen Maskierungsschwelle geformt werden, um so den wahrnehmbaren Nachhall zu unterdrücken [114, 120].

Geformte Impulsantworten innerhalb des Abhörvolumens Für eine globale Impulsantwort $g_{\boldsymbol{x}}(t)$ innerhalb des Abhörvolumens gilt $g_{\boldsymbol{x}}(t) = g_{\boldsymbol{l}}(t)$, falls $\boldsymbol{x}$ exakt einer Mikrofonposition $\boldsymbol{l}$ entspricht; andernfalls kann $g_{\boldsymbol{x}}(t)$ wie in (3.92) beschrieben aus den globalen Impulsantworten an den Mikrofonpositionen interpoliert werden. Es gilt

$$\begin{aligned} \mathrm{E}\{g_{\boldsymbol{x}}(t)\} &= \sum_{\boldsymbol{l}} \phi(\boldsymbol{x}, \boldsymbol{l}) \cdot \mathrm{E}\{g_{\boldsymbol{l}}(t)\} \\ &= \mathrm{const}(t) \cdot \sum_{\boldsymbol{l}} \phi(\boldsymbol{x}, \boldsymbol{l})\,, \end{aligned} \tag{3.97}$$

wobei $\phi(\boldsymbol{x}, \boldsymbol{l})$ in (3.90) gegeben ist.

Mit $\boldsymbol{k} = [k_x, k_y, k_z]^T$ gilt für die Einhüllende der quadrierten globalen Impulsantwort an der Position $\boldsymbol{x}$

$$\mathrm{E}\big\{|g_{\boldsymbol{x}}(t)|^2\big\} = \sum_{\boldsymbol{k}} r_g(t, \boldsymbol{k}) \cdot \Psi(\boldsymbol{x}, \boldsymbol{k})\,, \tag{3.98}$$

wobei

$$\Psi(\boldsymbol{x}, \boldsymbol{k}) = \sum_{\boldsymbol{l}} \phi(\boldsymbol{x}, \boldsymbol{l}) \cdot \phi(\boldsymbol{x}, \boldsymbol{l} + \boldsymbol{k})\,. \tag{3.99}$$

Für den Interpolationskern $\phi(\cdot)$ gelten die Eigenschaften

$$\sum_{\boldsymbol{l}} \phi(\boldsymbol{x}, \boldsymbol{l}) = 1, \tag{3.100}$$

sowie

$$\Psi(\boldsymbol{x}, \boldsymbol{k}) = \begin{cases} 1, & \text{für } \boldsymbol{k} = \boldsymbol{0}, \\ 0, & \text{sonst.} \end{cases} \tag{3.101}$$

Mit (3.100) und (3.101) gilt für die globale Impulsantwort an einer Position $\boldsymbol{x}$, beschrieben durch (3.97), der Zusammenhang

$$\mathrm{E}\{g_{\boldsymbol{x}}(t)\} = \mathrm{const}(t) \tag{3.102}$$

sowie

$$\mathrm{E}\big\{g_{\boldsymbol{x}}^2(t)\big\} = r_g(t, \boldsymbol{0}) \leq \mathrm{E}\big\{g_0^2(t)\big\}\,, \quad \text{für } \; t_1 < t < t_2. \tag{3.103}$$

Gleichung (3.103) besagt, dass das Abklingverhalten der globalen Impulsantwort $g_{\boldsymbol{x}}(t)$ an einem beliebigen Punkt $\boldsymbol{x} = [x, y, z]^T$ innerhalb des Abhörvolumens im Mittel von der gleichen oberen Schranke $g_0(t)$ beschränkt wird, wie die geformten Impulsantworten an den Mikrofonpositionen.

Anders formuliert: Wenn das Abhörvolumen ausreichend dicht abgetastet wurde, kann von der Formung der Impulsantworten an den Messpositionen auf die Formung *aller* Raumimpulsantworten innerhalb des Volumens geschlossen werden.

Im Verlauf dieser Arbeit konnte das in [114] erarbeitete Verfahren anhand gemessener Raumimpulsantworten validiert werden [78].

3.4.6 Weitere Ansätze zur Erhöhung der Robustheit gegenüber Fehlplatzierung

Neben der Abtastung des Abhörvolumens durch multiple Messungen von Raumimpulsantworten und der gleichzeitigen Formung aller gemessenen Raumimpulsantworten existieren weitere Ansätze zur Erhöhung der Robustheit gegenüber Fehlplatzierungen.

Im Folgenden werden weitere Ansätze aus der Literatur zur Erhöhung der räumlichen Robustheit vorgestellt.

Statistische Ansätze Wie bereits in der Einleitung zu diesem Unterabschnitt angemerkt kann die Veränderung des akustischen Übertragungskanals durch eine Fehlplatzie-

rung als additive Störung modelliert werden (siehe Abschnitt 3.4.1). In erster Näherung kann die Perturbation als unkorreliertes gaussförmiges Rauschen modelliert werden. Das sich aus dieser Annahme ergebende Optimalitätskriterium minimiert die Energie des Entzerrers [61, 88].

Die Erhöhung der räumlichen Robustheit der Entzerrer durch die Minimierung der Filterenergie konnte in [45] nachgewiesen werden.

In [82] wurden die Ergebnisse von Radlović et al. [139, 140] in den Filterentwurf integriert. Das von Kallinger et al. in [82] aufgestellte Optimierungsproblem gleicht dem in Gleichung (3.79); in [82] wurde jedoch das Problem der Übersprechkompensation (siehe Abschnitt 3.5) betrachtet.

Frequenzselektive Entzerrung Wie in dem Fehlermodell von Radlović et al. [139, 140] ersichtlich ist, nimmt die räumliche Robustheit zu hohen Frequenzen hin stark ab. Kolundžija et al. empfehlen in [96] eine Entzerrung lediglich bis zu einer Frequenz von $200\,\mathrm{Hz}$ vorzunehmen. Für die Zeitbereichsdarstellung der Filter fordern Kolundžija et al. ein der temporalen Maskierungsschwelle des menschlichen Hörapparates entsprechendes Abklingen und sehr kurze Entzerrer. Die Motivation hierfür ist, dass ein entsprechender Filter im Falle einer Fehlplatzierung zu keinem zusätzlich wahrnehmbaren Nachhall führt [96].

Für die Übersprechkompensation (siehe Abschnitt 3.5) wurden ähnliche Ergebnisse von Cooper et al. [18] gefunden. So empfehlen Cooper et al., keine Übersprechkompensation für hohe Frequenzen durchzuführen.

Kurze Entzerrer Von Kolundžija et al. [96] sowie von Kodrasi et al. [91] wurde der Zusammenhang zwischen der Filterlänge und der Robustheit gegenüber Fehlschätzungen der tatsächlichen Raumimpulsantwort untersucht.

Intuitiv ist zu erwarten, dass ein *kurzer* Entzerrer das Abklingverhalten im Falle von Fehlschätzungen weniger stark beeinflusst als ein *langes* Filter. Eine ähnliche Argumentation kann für den Betragsfrequenzgang des globalen Systems angenommen werden - ein kurzes Filter kann durch Fensterung im Zeitbereich aus einem langen Filter extrahiert werden. Das Äquivalent zur Fensterung im Zeitbereich ist die Faltung im Frequenzbereich, also üblicherweise eine Tiefpassfilterung [119] und damit eine Glättung des Betragsfrequenzganges.

In [96] und [91] wurden diese Annahmen bestätigt: Es wurde nachgewiesen, dass kürzere Entzerrer robuster als lange Entzerrer sind.

Adaptive Verfahren Mit der gesteigerten Rechenleistung aktueller Hardware ist eine adaptive Entzerrung möglich. In [152] wird eine Kamera verwendet, um die Kopfposition des Zuhörers relativ zu den Lautsprechern zu erfassen. Aufgrund der Kopfposition wird die Verzögerung der Direktschallkomponente von den Lautsprechern zu den Ohren geschätzt. Die Vorfilter für die Übersprechkompensation (siehe Abschnitt 3.5) werden auf Grundlage dieser Schätzung berechnet.

In [151] und [150] erweitern Song et al. das Verfahren aus [152] um die Modellierung der frühen Reflexionen (siehe Abschnitt 2.1.3). Hierzu wird auf Grundlage der Ausrichtung und Position des Kopfes des Zuhörers, sowie auf Grundlage der Position innerhalb des Raumes und der Raumdimensionen mit Hilfe der Spiegelmethode [2] der frühe Teil der Raumimpulsantwort simuliert.

Die Vorfilter für die Übersprechkompensation werden auf Grundlage der simulierten akustischen Kanäle entworfen. In [150] konnte experimentell gezeigt werden, dass mit diesem Verfahren die Ortbarkeit virtueller Schallquellen gesteigert werden kann.

3.4.7 Zusammenfassung

In diesem Abschnitt wurden die in der Literatur üblicherweise verwendeten Verfahren zum Entwurf von robusten Entzerrern vorgestellt. Hierbei wird (je nach Quelle) allgemein von *Robustheit*, von Robustheit gegenüber Kanalfehlschätzungen oder auch von Robustheit gegenüber räumlicher Fehlplatzierung gesprochen.

Lediglich Mei et al. [114] und Kolundžija et al. [96] widmen sich explizit dem Problem der räumlichen Robustheit. In [114] wird durch multiple Messungen von Raumimpulsantworten das Abklingverhalten der Raumimpulsantworten an jedem Punkt innerhalb eines begrenzten Volumens geformt. In [96] wird lediglich das Abklingverhalten des Entzerrers vorgegeben. Kolundžija et al. argumentieren, dass ein *kurzer*, im Zeitbereich *schnell* abklingender Entzerrer auch im Falle von räumlicher Fehlplatzierung zu keinem zusätzlichen Nachhall in der globalen Impulsantwort führen sollte.

Im Allgemeinen liefert die Entzerrung auf Grundlage vieler gemessener Raumimpulsantworten die *besten* Ergebnisse. Der hohe Aufwand für die Messung der benötigten Raumimpulsantworten ist der größte Nachteil dieses Verfahrens [79].

Radlović et al. untersuchten in [140] und [139] das Verhalten von einem *perfekten* Entzerrer im Fall von räumlicher Fehlplatzierung; hierbei wurde lediglich der Betragsfrequenzgang betrachtet. Radlović et al. kommen zu dem Schluss, dass im Falle von räumlicher Fehlplatzierung gänzlich von einer Entzerrung abgesehen werden sollte.

Ein weiteres in diesem Abschnitt betrachtetes Verfahren zur robusten Übersprechkompensation basiert auf einer Echtzeit-Schätzung der Direktschallkomponente sowie der frühen Reflexionen der Raumimpulsantworten [150–152]; der (späte) Nachhall wurde dabei nicht modelliert. Als Gütekriterium wurde die Ortbarkeit virtueller Quellen durch einen menschlichen Zuhörer angesetzt.

Ein oft verwendeter Ansatz zum Entwurf robuster Entzerrer ist, die Energie der Filter als Regularisierungsterm in die Zielfunktion zu integrieren [61–63, 88, 95]. Durch die Beschränkung der Energie der Filter werden *starke* Anhebungen im Frequenzgang des Entzerrers vermieden.

Kallinger et al. haben in [82] die von Radlović et al. in [140] und [139] erarbeitete Fehlerbeschreibung in den Entwurf der Entzerrer integriert.

Abgrenzung zu den in dieser Arbeit entwickelten Verfahren Im Rahmen dieser Arbeit wurde eine breitbandige Entzerrung im Frequenzbereich sowie eine gleichmäßige Formung im Zeitbereich angestrebt; beide Kriterien sollen auch im Falle von räumlicher Fehlplatzierung erfüllt sein.

Die in der Literatur behandelten robusten Verfahren bieten keine Lösung für die gemeinsame Formung und Entzerrung eines akustischen Kanals. Sie erfüllen somit nicht die Ansprüche, die an die im Rahmen dieser Arbeit entwickelten Methoden gestellt werden.

3.5 Übersprechkompensation

Das Problem der Übersprechkompensation ist mit dem der Kanalkompensation verwandt. Der Zusammenhang wird im Rahmen dieses Abschnitts aufgezeigt. Abbildung 3.22 stellt das in der Literatur üblicherweise betrachtete Szenario zur Übersprechkompensation in akustischen Szenarien dar: Ein binaurales Audiosignal soll einem Zuhörer über zwei Lautsprecher dargeboten werden.

Eine Darstellung der Signalpfade des in Abbildung 3.22 dargestellten Szenarios ist in Abbildung 3.23 gegeben. Die Quellensignale $s_1(n)$ und $s_2(n)$ sollen über ein Filternetzwerk, bestehend aus den Kompensationsfiltern $h_{\ell q}(n)$, $\ell, q \in \{1, 2\}$, und dem Mischsystem $c_{m\ell}(n)$, $m, \ell \in \{1, 2\}$, so übertragen werden, dass an der ersten Signalsenke $y_1(n) = s_1(n)$ gilt und an der zweiten Signalsenke $y_2(n) = s_2(n)$.

Der Entwurf eines Filternetzwerks zur Übersprechkompensation auf Grundlage eines gegebenen Mischsystems ist mit dem Problem der (blinden) Quellentrennung verwandt.

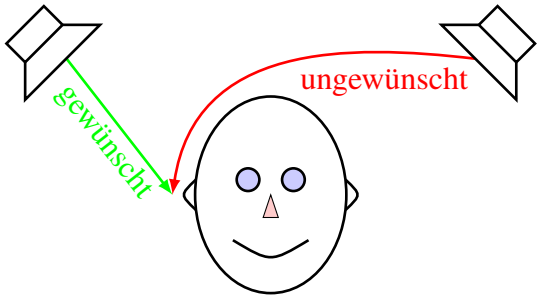

Abbildung 3.22: Übersprechen in einem akustischen Szenario. Ein binaurales Audiosignal wird über zwei Lautsprecher dargeboten. Das Ohr des Zuhörers wird von einem *gewünschten* beziehungsweise lateralen Signalanteil (grün), aber auch von einem *ungewünschten*, ipsilateralen Signalanteil (rot), der Übersprechkomponente, erreicht.

Während bei der blinden Quellentrennung lediglich die aufgenommenen Signale zur Verfügung stehen, um das Entmischsystem zu schätzen [109, 112], liegt bei den im Rahmen dieser Arbeit betrachteten Verfahren eine Messung beziehungsweise eine Schätzung des Mischsystems vor [79].

In Abschnitt 3.5.1 wird das übliche Vorgehen zum Filterentwurf für die Übersprechkompensation durch die Optimierung eines quadratischen Fehlerkriteriums vorgestellt. In Abschnitt 3.5.2 wird ein kurzer Überblick zu aktuellen Arbeiten und zur praktischen Nutzbarmachung der Übersprechkompensation in akustischen Szenarien gegeben.

Von praktischer Bedeutung ist die Übersprechkompensation in akustischen Szenarien beispielsweise für die binaurale Wiedergabe von Audiosignalen über Lautsprecher [21]. Während bei der Wiedergabe über Kopfhörer sichergestellt werden kann, dass jedes Ohr des Zuhörers nur von dem vorgesehenen Quellensignal erreicht wird, kann dies bei der Lautsprecherwiedergabe nicht garantiert werden (siehe Abbildung 3.22).

Beschreibung der Übertragungskanäle Ausgehend von dem in Abbildung 3.23 dargestellten Szenario lassen sich die Ausgangssignale $y_1(n)$ und $y_2(n)$ des Gesamtsystems als

$$\begin{aligned} y_1(n) &= s_1(n) * \overbrace{(h_{11}(n) * c_{11}(n) + h_{21}(n) * c_{12}(n))}^{g_{11}(n)} + \\ &\quad s_2(n) * \underbrace{(h_{12}(n) * c_{11}(n) + h_{22}(n) * c_{12}(n))}_{g_{12}(n)} \end{aligned} \tag{3.104}$$

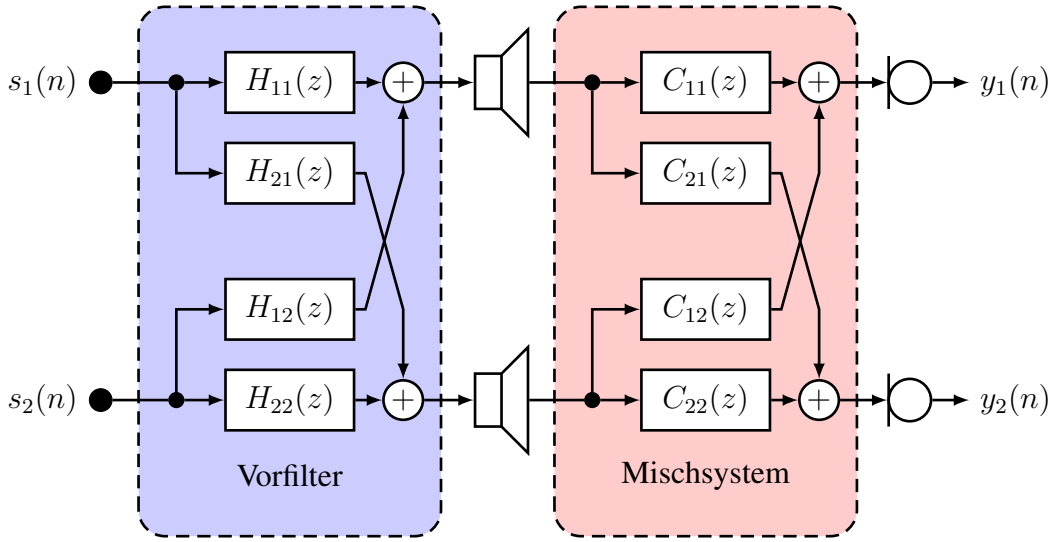

Abbildung 3.23: Darstellung des üblicherweise in der Literatur verwendeten 2×2 Szenarios zur Übersprechkompensation. Zwei Quellensignale werden über ein Filternetzwerk zu zwei Signalsenken übertragen.

und

$$\begin{aligned} y_2(n) &= s_1(n) * \overbrace{(h_{11}(n) * c_{21}(n) + h_{21}(n) * c_{22}(n))}^{g_{21}(n)} + \\ &\quad s_2(n) * \underbrace{(h_{12}(n) * c_{21}(n) + h_{22}(n) * c_{22}(n))}_{g_{22}(n)} \end{aligned} \tag{3.105}$$

beschreiben.

Aus der Forderung, dass $y_1(n) = s_1(n)$ sowie $y_2(n) = s_2(n)$ lassen sich die Übertragungskanäle $g_{mq}(n)$, $m, q \in \{1, 2\}$, in ipsilaterale (*gewünschte*) Pfade und kontralaterale (*ungewünschte*) Signalpfade einteilen. Es gilt also die Vorfilter $h_{\ell q}(n)$ derart zu entwerfen, dass die gewünschten Pfade $g_{11}(n)$ und $g_{22}(n)$ keinen wahrnehmbaren Einfluss aus das jeweilige Quellensignal haben. Die Signalpfade der Übersprechkomponenten $g_{12}(n)$ und $g_{21}(n)$ sollen idealerweise gleich null sein.

3.5.1 Quadratisches Optimalitätskriterium für den Entwurf der Filter zur Übersprechkompensation

Das mathematische Vorgehen für den Entwurf der Vorfilter zur Übersprechkompensation kann direkt aus dem in Abschnitt 3.2.1.2 vorgestellten Verfahren zur *Kanalkompensation mit gegebenem Zielsystem* abgeleitet werden.

Ausgehend von (3.104) und (3.105) können die einzelnen Übertragungskanäle in Vektor-Notation geschrieben werden:

$$\boldsymbol{g}_{11} = \boldsymbol{C}_{11}\boldsymbol{h}_{11} + \boldsymbol{C}_{12}\boldsymbol{h}_{21}, \tag{3.106}$$

$$\boldsymbol{g}_{12} = \boldsymbol{C}_{11}\boldsymbol{h}_{12} + \boldsymbol{C}_{12}\boldsymbol{h}_{22}, \tag{3.107}$$

$$\boldsymbol{g}_{21} = \boldsymbol{C}_{21}\boldsymbol{h}_{11} + \boldsymbol{C}_{22}\boldsymbol{h}_{21}, \tag{3.108}$$

$$\boldsymbol{g}_{22} = \boldsymbol{C}_{21}\boldsymbol{h}_{12} + \boldsymbol{C}_{22}\boldsymbol{h}_{22}, \tag{3.109}$$

wobei $\boldsymbol{C}_{m\ell}$ jeweils die Faltungsmatrix der individuellen akustischen Kanäle $c_{m\ell}(n)$ ist und $\boldsymbol{h}_{\ell q}$ die Filterkoeffizienten der Vorfilter $h_{\ell q}(n)$ enthält. Analog zu (3.18) wird für die gewünschten und für die ungewünschten Kanäle jeweils ein zu approximierendes Zielsystem vorgegeben. Für den gewünschten Signalpfad wird üblicherweise ein verzögerter und/oder bandpassgefilterter Einheitsimpuls angesetzt; der ungewünschte Signalpfad soll idealerweise konstant null sein [82].

Der Vektor $\boldsymbol{d}_d$ enthält die Koeffizienten des zu approximierenden Zielsystems $d_d(n)$für die gewünschten Signalpfade, während der Vektor $\boldsymbol{d}_c$ dem Nullvektor entspricht - das zu approximierende Zielsystem für die Übersprechpfade ist also konstant null:

$$\boldsymbol{d}_d = \left[d_d(0)\,, \ldots, d_d(L_g - 1)\right]^T, \tag{3.110}$$

$$\boldsymbol{d}_c = \boldsymbol{0} = \left[0, \ldots, 0\right]^T. \tag{3.111}$$

Mit

$$\boldsymbol{h} = \left[\boldsymbol{h}_{11}^T, \boldsymbol{h}_{12}^T, \boldsymbol{h}_{21}^T, \boldsymbol{h}_{22}^T\right]^T, \tag{3.112}$$

$$\boldsymbol{d} = \left[\boldsymbol{d}_d^T, \boldsymbol{d}_c^T, \boldsymbol{d}_c^T, \boldsymbol{d}_d^T\right]^T, \tag{3.113}$$

sowie

$$\boldsymbol{C} = \begin{bmatrix} \breve{\boldsymbol{C}} & \boldsymbol{0} \\ \boldsymbol{0} & \breve{\boldsymbol{C}} \end{bmatrix}, \qquad \breve{\boldsymbol{C}} = \begin{bmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{bmatrix} \tag{3.114}$$

kann das Optimierungsproblem aus (3.18) direkt für den Entwurf der Filter zur Übersprechkompensation verwendet werden:

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad : \quad \|\boldsymbol{C}\boldsymbol{h} - \boldsymbol{d}\|_2^2\,. \tag{3.115}$$

Die Lösung zu (3.115) kann durch die Moore-Penrose Pseudoinverse [134] der Kanalmatrix $\boldsymbol{C}$ in geschlossener Form angegeben werden:

$$\boldsymbol{h} = \boldsymbol{C}^{+}\boldsymbol{d}. \tag{3.116}$$

3.5.2 Weitere Ansätze zur Übersprechkompensation

In frühen Arbeiten zur binauralen Audioreproduktion über Lautsprecher wurden Freifeldbedingungen angenommen; es wurde also kein Nachhall betrachtet. Die akustischen Kanäle $c_{m\ell}(n)$ von den Lautsprechern zu den Ohren werden somit (je nach Modellierung) entweder lediglich durch die Laufzeitunterschiede oder durch die kopfbezogenen Übertragungsfunktionen (englisch *head-related transfer functions*, HRTFs) bestimmt [21]. Die kopfbezogenen Übertragungsfunktionen kodieren, wie ein aus einer bestimmten Richtung einfallendes Audiosignal durch Beugung, Brechung und Verdeckungseffekte verzerrt wird, bevor es die Innenohren erreicht. Kopfbezogene Übertragungsfunktionen werden auch zur Simulation virtueller akustischer Szenarien verwendet [100, 113]. Die Messung von HRTFs erfolgt entweder durch individualisierte Messungen mit Mikrofonen im Ohr des Zuhörers oder durch Messungen mit einem Kunstkopf [28, 142].

Die räumlich robuste Übersprechkompensation ist für die praktische Nutzbarmachung dieser Verfahren von großer Bedeutung. Im Allgemeinen lassen sich die Methoden zum Entwurf robuster Filter zur Kanalkompensation (siehe Abschnitt 3.4) auch für den Entwurf robuster Filter zur Übersprechkompensation anwenden. In [82] wurde das von Radlović et al. in [139] erarbeitete Fehlermodell in den Entwurf von Filtern zur Übersprechkompensation integriert. Ward schlug in [164] vor, die Filter zur Übersprechkompensation für viele Positionen gleichzeitig zu entwerfen, und so die räumliche Robustheit zu steigern.

In [165] wird darauf hingewiesen, dass die quantitative Bewertung der Qualität der Übersprechkompensation in akustischen Szenarien sehr schwer ist. So lässt sich die Ortbarkeit virtueller Quellen kaum durch quantitative Maße auf Grundlage der Übertragungskanäle erfassen, da überwiegend psychoakustische Effekte die Richtungswahrnehmung bestimmen [13, 165].

Gardner schlug in [38] vor, die Übersprechkompensation durch eine Verfolgung der Kopfposition und -ausrichtung räumlich robust zu gestalten. Aufgrund der Position und Ausrichtung des Kopfes des Zuhörers wird jeweils ein Satz an Filtern ausgewählt und als Vorfilternetzwerk verwendet. Die Beurteilung der Güte der Übersprechkompensation erfolgte durch Hörtests, in denen Testpersonen die wahrgenommene Richtung der virtu-

ellen Quelle angeben sollen. Eine Übersicht über das von Gardner entwickelte System zur räumlichen Darbietung von Audiosignalen wird in [37] gegeben.

Auch in den Arbeiten von Song et al. [150–152] wird von einer Verfolgung der Kopfposition ausgegangen. In [152] werden lediglich die Direktschallkomponenten für den Entwurf der Filter zur Übersprechkompensation betrachtet. In [151] und [150] werden mit Hilfe des Spiegelquellenmodells nach Allen und Berkley [2] zusätzlich zur Direktschallkomponente noch die frühen Reflexionen in den Übertragungskanälen geschätzt; der Entwurf der Filter zur Übersprechkompensation geschieht auf Grundlage der so geschätzten Übertragungskanäle. In Hörversuchen konnten Song et al. zeigen, dass die Modellierung der frühen Reflexionen für die Übersprechkompensation die Ortbarkeit virtueller Quellen in halligen Umgebungen steigern kann. Diese Ergebnisse konnten in einer im Institut für Signalverarbeitung an der Universität zu Lübeck angefertigten Bachelorarbeit reproduziert werden [5].

Das bereits in Abschnitt 3.2.2 vorgestellte Verfahren zur Kanalkompensation mit gegebenem Zielsystem durch die Minimierung der Maximumsnorm wurde von Betlehem et al. in [11] ursprünglich zur Übersprechkompensation vorgeschlagen. Betlehem et al. erweiterten hierzu das in Gleichung (3.115) gegebene Optimierungsproblem um eine zeitvariante Gewichtung des Fehlers und ersetzten die quadratische Norm durch eine Maximumsnorm.

3.6 Zusammenfassung

In diesem Kapitel wurden die wichtigsten Verfahren der kanalbasierten Raumhallkompensation vorgestellt. Es wird hierbei zwischen der Approximation eines gewünschten Zielsystems und der Formung der globalen Impulsantwort unterschieden. Des Weiteren wurde das Problem der robusten Kanalkompensation eingeführt. Es wurde vorgestellt, wie sich perturbierte Kanäle beschreiben lassen. Im Rahmen dieser Arbeit werden Verfahren entwickelt, die eine räumlich robuste Kanalkompensation ermöglichen. Um dieses Problem zu erfassen, wurden grundlegende Untersuchungen zur räumlichen Robustheit vorgestellt.

In einem weiteren Abschnitt wurde das Problem der Übersprechkompensation vorgestellt. Es wurde gezeigt, wie sich ein Verfahren zur kanalbasierten Raumhallkompensation mit gegebenem Zielsystem zum Entwurf von Filtern zur Übersprechkompensation erweitern lässt.

4 Erweiterte Methoden

Dieses Kapitel bildet den Kern dieser Dissertation. Hier werden die neu entwickelten Methoden zur Verbesserung der Qualität sowie zur Erhöhung der Robustheit gegenüber räumlicher Fehlplatzierung präsentiert. Dazu wird in Abschnitt 4.1 ein allgemeines Rahmenwerk zur Formung von Raumimpulsantworten im Zeitbereich entwickelt. In Abschnitt 4.2 werden zwei neue Verfahren zur Regularisierung im Frequenzbereich erarbeitet; diese integrieren die Forderung nach einem flachen, beziehungsweise beliebig geformten Betragsfrequenzgang in das Optimierungsproblem. In Abschnitt 4.3 werden zwei neue Verfahren entwickelt, um die Robustheit gegenüber räumlicher Fehlplatzierung zu erhöhen. Das Problem der Übersprechkompensation wird durch die Wahl der Gewichtsfenster für die Formung der globalen Impulsantworten gelöst und in Abschnitt 4.4.1 behandelt.

4.1 Mehrkanalige Formung der Impulsantworten im Zeitbereich

Für die Entwicklung der Algorithmen wird ein Formelwerk für den allgemeinen Fall mit Q Quellensignalen, N Lautsprechern und M Mikrofonen aufgestellt. Das hier entwickelte Rahmenwerk wurde in [79] publiziert. Der Aufbau eines solchen Systems ist exemplarisch in Abbildung 4.1 dargestellt.

In Abbildung 4.1 ist dargestellt, wie das Vorfilternetzwerk, bestehend aus den Filtern $H_{\ell q}(z)$, und das durch den Raum gegebene Mischsystem, bestehend aus den Filtern $C_{m\ell}(z)$, miteinander verschaltet sind; hierbei ist $1 \leq q \leq Q$ der Index des Quellensignals, $1 \leq \ell \leq N$ der Index des Lautsprechers und $1 \leq m \leq M$ der Index des Mikrofons. Für jede Mikrofonposition können des Weiteren mehrere Realisierungen der Raumimpulsantworten in den Filterentwurf integriert werden. Die zusätzlichen Realisierungen können beispielsweise durch R zusätzliche Messungen in der direkten Umgebung der Referenzposition gewonnen werden.

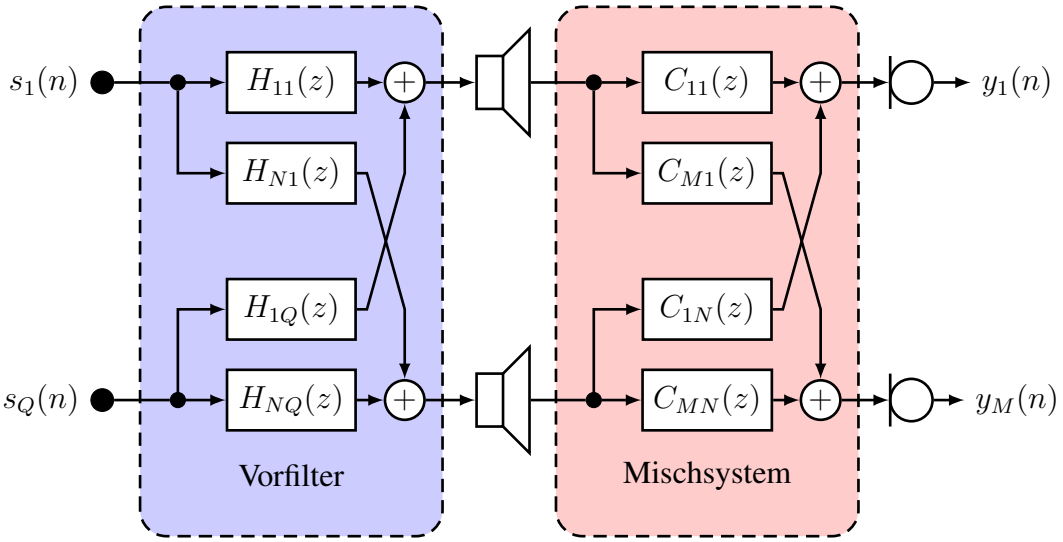

Abbildung 4.1: Allgemeines Szenario mit Q Quellensignalen, N Lautsprechern und M Mikrofonen. Für jede Referenzposition der Mikrofone können R Realisierungen der Raumimpulsantwort vorliegen (beispielsweise durch zusätzliche Messungen in der Nähe der Referenzposition).

Sowohl die Raumhallkompensation als auch die gemeinsame Raumhall- und Übersprechkompensation lassen sich in diesem Rahmenwerk formulieren.

Unter der Annahme, dass alle beteiligten Systeme – sowohl die Raumimpulsantworten $C_{m\ell}(z)$ als auch die Vorfilter $H_{\ell q}(z)$ – linear und zeitinvariant sind, lassen sich die globalen Impulsantworten von einer Signalquelle q zu einem Mikrofon m als

$$G_{mq}(z) = \sum_{\ell=1}^{N} C_{m\ell}(z)\, H_{\ell q}(z)\,, \quad q = 1, \ldots, Q, \;\; m = 1, \ldots, M, \tag{4.1}$$

berechnen.

Die globalen Impulsantworten können in *gewünschte* und *unerwünschte* (zum Beispiel Übersprechen) Kanäle gruppiert werden. Hierfür muss für jede Abhörposition m definiert werden, welches Quellensignal $s_q(n)$ als gewünscht und welches als unerwünschtes Übersprechen behandelt werden soll. Im Falle einer *idealen* Entzerrung und Übersprechkompensation würde

$$G_{mq}(z) = 0 \tag{4.2}$$

für die Übersprechkomponenten und

$$G_{mq}(z) = D_{mq}(z) \tag{4.3}$$

für die gewünschten Übertragungskanäle gelten; $D_{mq}(z)$ beschreibt hierbei ein gewünschtes Zielsystem.

Die Vorfilter $h_{\ell q}(n)$ werden unabhängig für jedes der Q Quellensignal entworfen. Die folgenden Herleitungen beziehen sich also ohne Beschränkung der Allgemeinheit auf die Annahme, dass jeweils nur ein Quellensignal aktiv ist.

Die einzelnen in Kapitel 3 vorgestellten Szenarien lassen sich in dem hier entwickelten Formelwerk folgendermaßen abbilden:

- **Einkanalige Entzerrung:** Für $Q = N = M = 1$ ergibt sich für die einzige globale Übertragungsfunktionen der Zusammenhang $G_{11}(z) = H_{11}(z)\, C_{11}(z)$. In diesem Szenario gilt es also, einen einzigen Entzerrer $H_{11}(z)$ für einen einzigen Kanal $C_{11}(z)$ zu optimieren.

- **Übersprechkompensation:** In der Literatur wird die Übersprechkompensation üblicherweise mit $Q = 2$ Quellensignalen, $N = 2$ Lautsprechern und $M = 2$ Abhörpositionen beschrieben [11, 150, 164]. Es ergeben sich somit insgesamt vier Übertragungskanäle, die in zwei gewünschte und zwei unerwünschte Kanäle gruppiert werden. Das hier entwickelte Rahmenwerk erlaubt eine Generalisierung auf beliebig viele Quellensignale, Übertragungskanäle und Abhörpositionen.

- **Robuste einkanalige Entzerrung:** Auch die robuste einkanalige Entzerrung wird durch $Q = M = 1$ abgebildet. Die Anzahl der Lautsprecher kann hier jedoch variieren. Im Gegensatz zu [114] wird in dem hier entwickelten Rahmenwerk von nur *einer* Mikrofonposition gesprochen. Die Abtastung des Volumens wird durch $1 \leq r \leq R$ Realisierungen der Impulsantwort $c_{11}^{(r)}(n)$ beschrieben.

In den folgenden Abschnitten wird das Formelwerk für die Formung im Zeitbereich hergeleitet [79]. Die Erweiterungen zur Übersprechkompensation [79, 80], zur robusten Formung [73, 76–80] sowie zur gleichzeitigen Entzerrung des Frequenzgangs durch entsprechende Regularisierungsterme [74, 75, 78, 79] werden in den späteren Abschnitten dieses Kapitels entwickelt.

4.1.1 Formulierung im Zeitbereich

Unter der Annahme, dass alle beteiligten Systeme linear und zeitinvariant sind, sowie eine endliche Impulsantwort besitzen, werden die Vektoren

$$\boldsymbol{h}_{\ell q} = \left[h_{\ell q}(0)\,, h_{\ell q}(1)\,, \ldots, h_{\ell q}(L_h - 1)\right]^T\,, \tag{4.4}$$

$$\boldsymbol{d}_{mq} = \left[d_{mq}(0)\,, d_{mq}(1)\,, \ldots, d_{mq}(L_g - 1)\right]^T\,, \tag{4.5}$$

definiert. Die Länge der Vorfilter wird mit L_h, die Länge der Raumimpulsantworten mit L_c und die Länge der Gesamtsysteme mit

$$L_g = L_h + L_c - 1 \tag{4.6}$$

beschrieben. Mit $\boldsymbol{C}_{m\ell}$ als Faltungsmatrix (3.6) der Raumimpulsantwort $c_{m\ell}(n)$ kann das allgemeine Problem der mehrkanaligen Kanalkompensation für ein Quellensignal $s_q(n)$ als

$$\boldsymbol{C}\boldsymbol{h}_q = \boldsymbol{d}_q \tag{4.7}$$

formuliert werden. Die Matrix $\boldsymbol{C}$ sowie die Vektoren $\boldsymbol{h}_q$ und $\boldsymbol{d}_q$ sind dabei als

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_{11} & \cdots & \boldsymbol{C}_{1N} \\ \vdots & \ddots & \vdots \\ \boldsymbol{C}_{M1} & \cdots & \boldsymbol{C}_{MN} \end{bmatrix}, \quad \boldsymbol{h}_q = \begin{bmatrix} \boldsymbol{h}_{1q} \\ \vdots \\ \boldsymbol{h}_{Nq} \end{bmatrix} \text{ und } \boldsymbol{d}_q \begin{bmatrix} \boldsymbol{d}_{1q} \\ \vdots \\ \boldsymbol{d}_{Mq} \end{bmatrix} \tag{4.8}$$

definiert. Ist das Signal $s_q(n)$ an der Mikrofonposition m als ungewünschtes Übersprechen definiert, so wird $\boldsymbol{d}_{mq}$ als Nullvektor der Länge L_g gewählt; andernfalls beschreibt $\boldsymbol{d}_{mq}$ ein gewünschtes Systemverhalten für den Übertragungskanal $g_{mq}(n)$.

Gleichung (4.7) kann als Generalisierung von (3.19) auf beliebig viele Quellensignale, Übertragungskanäle und Abhörpositionen verstanden werden.

Das in (4.7) dargestellte Gleichungssystem kann im Sinne des kleinsten quadratischen Fehlers mit Hilfe der Moore-Penrose Pseudoinversen [134] der Matrix $\boldsymbol{C}$ gelöst werden:

$$\boldsymbol{h}_q = \boldsymbol{C}^{+}\boldsymbol{d}_q. \tag{4.9}$$

Nach dem *multichannel inversion theorem* (MINT) ist bei einer mehrkanaligen Signalübertragung sogar eine vollständige Invertierung der Kanäle möglich [125]. Die

vollständige Entzerrung nach MINT ist in keinster Weise robust gegenüber Kanalfehlschätzungen [55].

Für einen robusten Filterentwurf werden die Unsicherheiten bezüglich der Raumimpulsantworten analog zu dem in (3.78) dargestellten Gleichungssystem modelliert. Jede Raumimpulsantwort wird hierfür als potentiell fehlerbehaftet angenommen und als

$$\hat{c}_{m\ell}(n) = c_{m\ell}(n) + p_{m\ell}(n) \tag{4.10}$$

beschrieben. Hierbei bezeichnet $c_{m\ell}(n)$ die gemessene Impulsantwort, $p_{m\ell}(n)$ ist ein additiver Störterm, der beispielsweise die Veränderung einer Raumimpulsantwort durch eine Bewegung des Mikrofons beschreibt. Gleichung (4.7) wird entsprechend umformuliert und es ergibt sich

$$(\boldsymbol{C} + \boldsymbol{P})\, \boldsymbol{h}_q = \boldsymbol{d}_q, \tag{4.11}$$

wobei die Matrix $\boldsymbol{P}$ analog zur Matrix $\boldsymbol{C}$ die einzelnen Faltungsmatrizen $\boldsymbol{P}_{m\ell}$ der Perturbationen $p_{m\ell}(n)$ enthält:

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{P}_{11} & \cdots & \boldsymbol{P}_{1N} \\ \vdots & \ddots & \vdots \\ \boldsymbol{P}_{M1} & \cdots & \boldsymbol{P}_{MN} \end{bmatrix}. \tag{4.12}$$

Eine weitere Möglichkeit zum Entwurf robuster Entzerrer basiert auf dem Ansatz, mehrere Messungen eines akustischen Kanals zu verwenden und so unter Beachtung des räumlichen Abtasttheorems [1] die Raumimpulsantworten an mehreren räumlichen Positionen gleichzeitig zu formen [114]. In dem hier entwickelten Formelwerk wird angenommen, dass für jede Mikrofonposition m insgesamt R Messungen durchgeführt werden. Die einzelnen Realisierungen der Raumimpulsantworten $c_{m\ell}(n)$ werden als

$$c_{m\ell}^{(r)}(n)\,, \quad r = 1, \ldots, R, \tag{4.13}$$

bezeichnet. Mit $\boldsymbol{C}_{m\ell}^{(r)}$ als Faltungsmatrix der Raumimpulsantwort $c_{m\ell}^{(r)}(n)$ wird die r-te Realsierung der Kanalmatrix aus (4.8) als

$$\boldsymbol{C}^{(r)} = \begin{bmatrix} \boldsymbol{C}_{11}^{(r)} & \cdots & \boldsymbol{C}_{1N}^{(r)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{C}_{M1}^{(r)} & \cdots & \boldsymbol{C}_{MN}^{(r)} \end{bmatrix} \tag{4.14}$$

definiert. Die gleichzeitige Entzerrung aller R Realisierungen der Raumimpulsantworten kann kurz als

$$\begin{bmatrix} \boldsymbol{C}^{(1)} \\ \vdots \\ \boldsymbol{C}^{(R)} \end{bmatrix} \boldsymbol{h}_q = \begin{bmatrix} \boldsymbol{d}_q \\ \vdots \\ \boldsymbol{d}_q \end{bmatrix} \tag{4.15}$$

geschrieben werden.

Im Rahmen dieser Arbeit wird das Prinzip der Formung von Impulsantworten ohne vorgegebenes Zielsystem verfolgt. Im Folgenden wird die auf der p-Norm basierende Zielfunktion aus [120] entsprechend generalisiert.

4.1.2 Entwurf der Filter für die Formung im Zeitbereich

Die zu optimierende Zielfunktion für die Formung der globalen Impulsantworten im Zeitbereich ist die direkte Generalisierung von (3.57) aus [120] auf Q Quellensignale, N Lautsprecher und M Abhörpositionen, wobei für jede der $N \cdot M$ Raumimpulsantworten potentiell R Realisierungen vorliegen können. Da für jedes der Q Quellensignale N Vorfilter $h_{\ell q}(n)$, $q = 1, \ldots, Q$, $\ell = 1, \ldots, N$, berechnet werden, ergeben sich insgesamt Q individuelle Optimierungsprobleme [79]. Die Betrachtung von R unterschiedlichen Realisierungen einer Raumimpulsantwort erlaubt die Integration von unterschiedlichen Verfahren zur Erhöhung der Robustheit gegenüber räumlicher Fehlplatzierung; entsprechende Verfahren werden später in diesem Kapitel eingeführt. Das hier entwickelte Formelwerk wurde in [79] publiziert. Eine grafische Darstellung zur Verdeutlichen der Nomenklatur ist in Abbildung 4.2 gegeben. In Abbildung 4.2 wird exemplarisch der Aufbau für ein Quellensignal ($Q = 1$), zwei Lautsprecher ($N = 2$), einer Abhörposition ($M = 1$) und insgesamt drei Realisierungen ($R = 3$) der Raumimpulsantwort dargestellt.

Für $N = M = R = 1$ ist das hier erarbeitete Formelwerk äquivalent zu dem in [120] vorgestellten Verfahren; es kann also als eine Generalisierung von diesem verstanden werden. Das von Mei et al. in [114] vorgestellte robuste Verfahren basiert auf der räumlichen Abtastung eines Volumens. In [114] wurde ein Aufbau mit $N = 13$ Lautsprechern und $R = 57$ Realisierungen der N akustischen Kanäle um eine ($M = 1$) Referenzposition simuliert. Im Gegensatz zu dem hier erarbeiteten Verfahren wird in [114] nur der ungewünschte Teil der akustischen Kanäle optimiert.

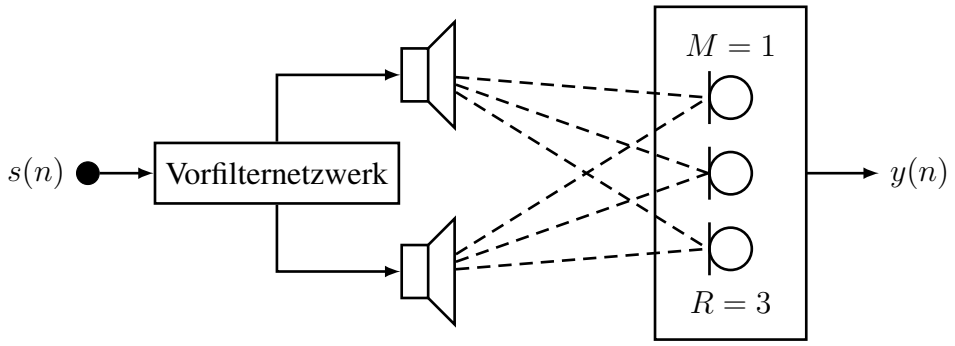

Abbildung 4.2: Grafische Darstellung des Aufbaus für $Q = 1$ Quellensignal mit $N = 2$ Lautsprechern, einer ($M = 1$) Referenzposition und insgesamt $R = 3$ Realisierungen des akustischen Kanals. Die $R = 3$ Realisierungen können beispielsweise durch $R - 1$ zusätzliche Messungen um die Referenzposition generiert werden.

4.1.2.1 Kostenfunktion für die Formung im Zeitbereich

Mit Hilfe von zwei Gewichtsfenstern $w_{mq}^{(d)}(n)$ und $w_{mq}^{(u)}(n)$ werden die gewünschten und ungewünschten Teile der globalen Impulsantworten definiert. Der gewünschte (englisch *desired*) Teil der r-ten Realsierung der globalen Impulsantwort von der Quelle q zur Referenzposition m wird als

$$g_{mq}^{(r,d)}(n) = w_{mq}^{(d)}(n)\, g_{mq}^{(r)}(n)\,, \tag{4.16}$$

und der ungewünschte (englisch *unwanted*) Teil als

$$g_{mq}^{(r,u)}(n) = w_{mq}^{(u)}(n)\, g_{mq}^{(r)}(n)\,, \tag{4.17}$$

bestimmt.

Mit der auf der p-Norm basierenden Zielfunktion (3.57) aus [120] als Grundlage für die Formung im Zeitbereich ergeben sich Q individuelle Optimierungsprobleme der Form

$$\underset{\boldsymbol{h}_q}{\text{minimize}} \;:\; f_q(\boldsymbol{h}_q)\,, \;\; \text{s.t.} \;\; \|\boldsymbol{h}_q\|_2 = 1, \quad q = 1, \ldots, Q, \tag{4.18}$$

mit

$$f_q(\boldsymbol{h}_q) = \log\left(\frac{f_q^{(u)}(\boldsymbol{h}_q)}{f_q^{(d)}(\boldsymbol{h}_q)}\right). \tag{4.19}$$

Die Funktionen $f_q^{(d)}(\boldsymbol{h}_q)$ und $f_q^{(u)}(\boldsymbol{h}_q)$ sind als

$$f_q^{(d)}(\boldsymbol{h}_q) = \left\|\boldsymbol{g}_q^{(d)}\right\|_{p_d} \tag{4.20}$$

und

$$f_q^{(u)}(\boldsymbol{h}_q) = \left\|\boldsymbol{g}_q^{(u)}\right\|_{p_u} \tag{4.21}$$

definiert. Die Vektoren $\boldsymbol{g}_q^{(d)}$ und $\boldsymbol{g}_q^{(u)}$ enthalten hierbei die R Realisierungen der gewünschten (beziehungsweise ungewünschten) Komponenten aller globalen Impulsantworten:

$$\boldsymbol{g}_q^{(d)} = \left[\boldsymbol{g}_{1q}^{(1,d)^T}, \ldots, \boldsymbol{g}_{Mq}^{(1,d)^T}, \ldots, \boldsymbol{g}_{1q}^{(R,d)^T}, \ldots, \boldsymbol{g}_{Mq}^{(R,d)^T}\right]^T \tag{4.22}$$

sowie

$$\boldsymbol{g}_q^{(u)} = \left[\boldsymbol{g}_{1q}^{(1,u)^T}, \ldots, \boldsymbol{g}_{Mq}^{(1,u)^T}, \ldots, \boldsymbol{g}_{1q}^{(R,u)^T}, \ldots, \boldsymbol{g}_{Mq}^{(R,u)^T}\right]^T . \tag{4.23}$$

Die Vektoren $\boldsymbol{g}_q^{(d)}$ und $\boldsymbol{g}_q^{(u)}$ enthalten also jeweils $M \cdot R$ Impulsantworten der Länge L_g.

Das Optimierungsproblem (4.18) wird durch ein Gradientenabstiegsverfahren gelöst. Die Adaptationsregel für die Formung im Zeitbereich lautet

$$\boldsymbol{h}_q^{l+1} = \boldsymbol{h}_q^l - \mu^l \left(\nabla_{\boldsymbol{h}_q} f_q\left(\boldsymbol{h}_q^l\right)\right) , \tag{4.24}$$

wobei l den Iterationsschritt bezeichnet und μ^l die adaptive Schrittweite im l-ten Iterationsschritt ist. Die Nebenbedingung in (4.18) wird durch eine Normierung des Zielvektors $\boldsymbol{h}_q$ nach jedem Iterationsschritt hergestellt und verbessert die Stabilität der Optimierung.

4.1.2.2 Herleitung der Gradienten für die Formung im Zeitbereich

Die Optimierung von (4.18) erfolgt iterativ durch ein Gradientenabstiegsverfahren. Durch die Anwendung der Logarithmengesetze auf (4.19) lässt sich der Gradient bezüglich des Zielvektors $\boldsymbol{h}_q$ als

$$\nabla_{\boldsymbol{h}_q} f_q(\boldsymbol{h}_q) = \frac{1}{f_q^{(u)}(\boldsymbol{h}_q)} \nabla_{\boldsymbol{h}_q} f_q^{(u)}(\boldsymbol{h}_q) - \frac{1}{f_q^{(d)}(\boldsymbol{h}_q)} \nabla_{\boldsymbol{h}_q} f_q^{(d)}(\boldsymbol{h}_q) \tag{4.25}$$

ausdrücken. Die Berechnung des Gradienten $\nabla_{\boldsymbol{h}_q} f_q(\boldsymbol{h}_q)$ erfolgt durch die getrennte Berechnung der Gradienten $\nabla_{\boldsymbol{h}_q} f_q^{(d)}(\boldsymbol{h}_q)$ und $\nabla_{\boldsymbol{h}_q} f_q^{(u)}(\boldsymbol{h}_q)$.

Der Gradient für $f_q^{(d)}(\boldsymbol{h}_q)$ bezüglich des Zielvektors $\boldsymbol{h}_q$ wird durch

$$\nabla_{\boldsymbol{h}_q} f_q^{(d)}(\boldsymbol{h}_q) = \zeta_q^{(d)}(\boldsymbol{h}_q) \cdot \nabla_{\boldsymbol{h}_q} \phi_{f_q^{(d)}}(\boldsymbol{h}_q) \tag{4.26}$$

berechnet, wobei $\zeta_q^{(d)}(\boldsymbol{h}_q)$ als

$$\zeta_q^{(d)}(\boldsymbol{h}_q) = \left(\sum_{m=1}^{M} \sum_{r=1}^{R} \sum_{n=0}^{L_g-1} \left| g_{mq}^{(r,d)}(n) \right|^{p_d} \right)^{\frac{1}{p_d}-1} \tag{4.27}$$

und $\nabla_{\boldsymbol{h}_q} \phi_{f_q^{(d)}}(\boldsymbol{h}_q)$ durch

$$\nabla_{\boldsymbol{h}_q} \phi_{f_q^{(d)}}(\boldsymbol{h}_q) = \begin{bmatrix} \sum_{m=1}^{M} \sum_{r=1}^{R} \left(\boldsymbol{C}_{m1}^{(r)} \right)^T \boldsymbol{b}_{mq}^{(r,d)} \\ \vdots \\ \sum_{m=1}^{M} \sum_{r=1}^{R} \left(\boldsymbol{C}_{mN}^{(r)} \right)^T \boldsymbol{b}_{mq}^{(r,d)} \end{bmatrix} \tag{4.28}$$

mit $\boldsymbol{b}_{mq}^{(r,d)}$ als

$$\boldsymbol{b}_{mq}^{(r,d)} = \operatorname{diag}\left\{ \operatorname{sign}\left\{ \boldsymbol{g}_{mq}^{(r,d)} \right\} \right\} \operatorname{diag}\left\{ \boldsymbol{w}_{mq}^{(d)} \right\} \cdot \left| \boldsymbol{g}_{mq}^{(r,d)} \right|^{p_d-1} \tag{4.29}$$

berechnet wird [79].

Analog zu der Berechnung des Gradienten für den gewünschten Teil erfolgt die Berechnung des Gradienten für den ungewünschten Teil der globalen Impulsantworten $\nabla_{\boldsymbol{h}_q} f_q^{(u)}(\boldsymbol{h}_q)$. Der Gradient bezüglich des Zielvektors $\boldsymbol{h}_q$ wird durch

$$\nabla_{\boldsymbol{h}_q} f_q^{(u)}(\boldsymbol{h}_q) = \zeta_q^{(u)}(\boldsymbol{h}_q) \cdot \nabla_{\boldsymbol{h}_q} \phi_{f_q^{(u)}}(\boldsymbol{h}_q) \tag{4.30}$$

berechnet, wobei $\zeta_q^{(u)}(\boldsymbol{h}_q)$ als

$$\zeta_q^{(u)}(\boldsymbol{h}_q) = \left(\sum_{m=1}^{M} \sum_{r=1}^{R} \sum_{n=0}^{L_g-1} \left| g_{mq}^{(r,u)}(n) \right|^{p_u} \right)^{\frac{1}{p_u}-1} \tag{4.31}$$

und $\nabla_{\boldsymbol{h}_q} \phi_{f_q^{(u)}}(\boldsymbol{h}_q)$ durch

$$\nabla_{\boldsymbol{h}_q} \phi_{f_q^{(u)}}(\boldsymbol{h}_q) = \begin{bmatrix} \sum_{m=1}^{M} \sum_{r=1}^{R} \left(\boldsymbol{C}_{m1}^{(r)} \right)^T \boldsymbol{b}_{mq}^{(r,u)} \\ \vdots \\ \sum_{m=1}^{M} \sum_{r=1}^{R} \left(\boldsymbol{C}_{mN}^{(r)} \right)^T \boldsymbol{b}_{mq}^{(r,u)} \end{bmatrix} \tag{4.32}$$

mit $\boldsymbol{b}_{mq}^{(r,u)}$ als

$$\boldsymbol{b}_{mq}^{(r,u)} = \operatorname{diag}\{\operatorname{sign}\{\boldsymbol{g}_{mq}^{(r,u)}\}\} \operatorname{diag}\{\boldsymbol{w}_{mq}^{(u)}\} \cdot \left|\boldsymbol{g}_{mq}^{(r,u)}\right|^{p_u-1} \tag{4.33}$$

berechnet wird [79].

4.2 Regularisierung im Spektralbereich

Der von Mertins et al. in [120] vorgeschlagene Algorithmus zur Formung von Raumimpulsantworten präferiert üblicherweise Lösungen mit einem einzigen dominanten Zeitkoeffizienten im gewünschten Teil der globalen Impulsantwort. Damit weisen die geformten Impulsantworten im Allgemeinen einen flachen Frequenzgang auf.

In Abbildung 4.3 (untere Zeile) ist das Ergebnis der Formung einer gemessenen Raumimpulsantwort nach [120] mit langer Nachhallzeit dargestellt; deutlich sind die breitbandigen Störungen im Betragsfrequenzgang zu erkennen.

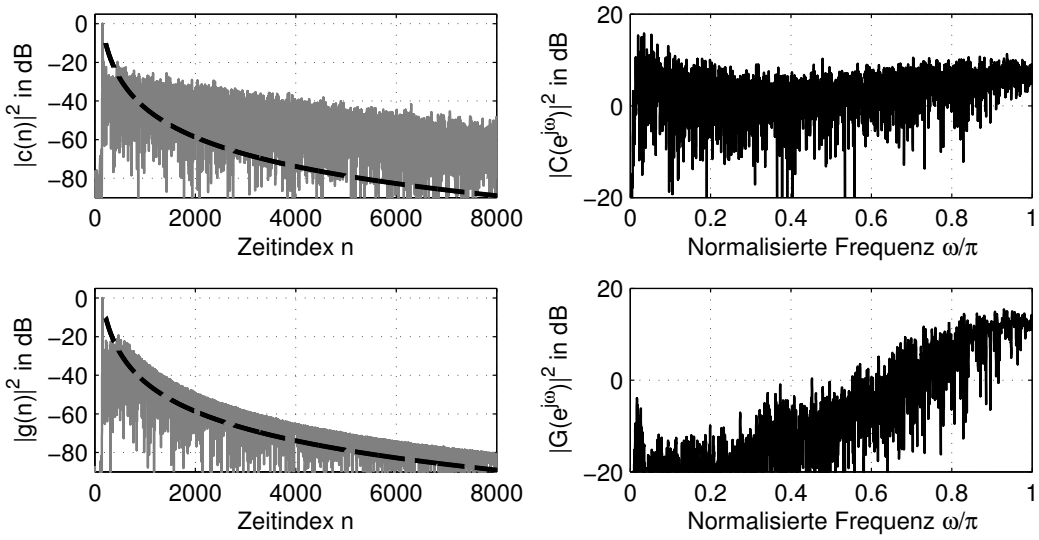

Abbildung 4.3: Oben: Gemessene Raumimpulsantwort mit langer Nachhallzeit im Zeitbereich sowie der Betragsfrequenzgang. Unten: Die nach [120] geformte Impulsantwort im Zeitbereich sowie der Betragsfrequenzgang.

Von Kallinger et al. wurde empfohlen, die spektralen Verzerrungen, die sich durch die alleinige Optimierung der Zeitbereichsdarstellung ergeben, mit einem zusätzlichen Entzerrer zu kompensieren [85]. Hierzu wurde in [85] ein lineares Prädiktorfehlerfilter verwendet. Das Prädiktorfilter der Schrittweite eins wurde für die globale Impulsant-

wort $g(n)$ entworfen; die Länge des Prädiktorfehlerfilters wurde *kurz* gewählt, um die Formung im Zeitbereich möglichst wenig zu beeinflussen [85].

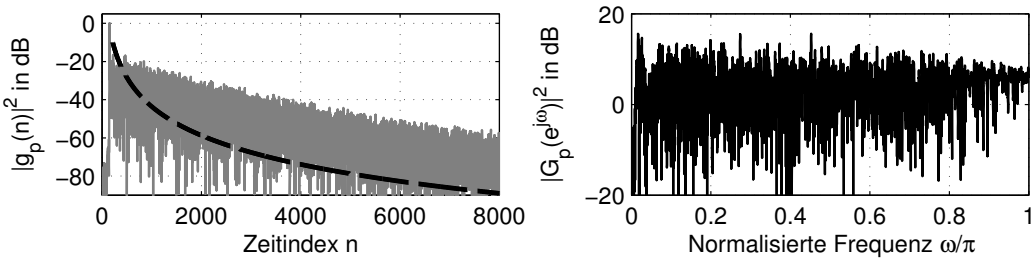

Abbildung 4.4: Die geformte Impulsantwort aus Abbildung 4.3 (untere Zeile) wurde mit der Postfilter-Methode nach Kallinger et al. [85] entzerrt. Das Gesamtsystem nach der Nachfilterung ($g_p(n)$) ist im Zeit- und im Frequenzbereich dargestellt.

Wird die in Abbildung 4.3 (untere Zeile) dargestellte, bereits geformte Impulsantwort mit der Postfilter-Methode nach Kallinger et al. [85] nachbearbeitet, so ergibt sich das in Abbildung 4.4 dargestellte Ergebnis. Die Länge des verwendeten Prädiktorfehlerfilters beträgt hierbei $L_p = 15$ Zeitkoeffizienten. Beim Vergleich von Abbildung 4.4 mit der gemessenen Raumimpulsantwort sowie der geformten Gesamtimpulsantwort (Abbildung 4.3, obere beziehungsweise untere Zeile) ist zu erkennen, dass durch die Nachfilterung nach Kallinger et al. [85] die erfolgreiche Formung im Zeitbereich nahezu vollständig zerstört wird.

Im Rahmen dieser Arbeit werden Methoden zur gleichzeitigen Optimierung der Zeit- sowie der Frequenzbereichsdarstellung der globalen Impulsantwort entwickelt. Hierfür wird ein Regularisierungsterm formuliert und in die Zielfunktion integriert. Das Verfahren wurde erstmalig in [81] publiziert. Der erarbeitete Regularisierungsterm optimiert die p-Norm des Betragsfrequenzgangs der globalen Impulsantwort (beziehungsweise des Entzerrers selbst). Das Verfahren wurde auf beliebig viele Übertragungskanäle erweitert und in [78] sowie [79] publiziert.

Eine alternative Möglichkeit zur Regularisierung des Frequenzgangs basiert auf einer generalisierten Form des *spectral flatness measure* (SFM, [51, 70, 105, 106]). Ein entsprechendes Optimalitätskriterium wurde im Rahmen dieser Arbeit aufgestellt, in die Kostenfunktion zum Filterentwurf integriert und in [75] sowie [74] publiziert.

Sowohl das auf der p-Norm als auch das auf dem SFM basierende Verfahren werden in diesem Abschnitt motiviert und entwickelt.

4.2.1 Regularisierung auf Basis der p-Norm

Im Verlauf dieser Arbeit wurde gezeigt, dass es Szenarien gibt, in denen das auf der p-Norm basierende Verfahren zur Formung von Raumimpulsantworten nach [120] zu starken Verzerrungen des Betragsfrequenzgangs des Gesamtsystems führt [81]. Eine nachträgliche Entzerrung des Frequenzgangs nach der Methode von Kallinger et al. [85] führt in vielen Fällen zu einem Gesamtsystem mit ähnlichen akustischen Eigenschaften wie die nicht geformte Raumimpulsantwort (siehe Abbildungen 4.3 sowie 4.4).

Im Rahmen dieser Arbeit wurde ein Verfahren erarbeitet, welches die p-Norm des Betragsfrequenzgangs als zusätzliches Optimalitätskriterium in den Filterentwurf integriert. Durch die Minimierung der p-Norm mit einem ausreichend großen Wert für p (üblicherweise $p \approx 8$) werden starke Anhebungen im Betragsfrequenzgang unterdrückt. Es ergeben sich zwei mögliche Anwendungen:

1. Der Betragsfrequenzgang des Entzerrers wird optimiert.
2. Der Betragsfrequenzgang des Gesamtsystems wird optimiert.

Beide Verfahren wurden im Rahmen dieser Arbeit entwickelt und evaluiert; sie wurden anschließend in [81] publiziert. Das erweiterte Optimierungsproblem lautet

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad : \quad f(\boldsymbol{h}) + \alpha y(\boldsymbol{h}) \,, \quad \text{s.t.} \quad \|\boldsymbol{h}\|_2 = 1, \tag{4.34}$$

wobei $f(\boldsymbol{h})$ die auf der p-Norm basierende Zielfunktion aus [120] (siehe Gleichung (3.57)), $\alpha \geq 0$ ein Gewichtsfaktor ist und $y(\boldsymbol{h})$ der zusätzliche Regularisierungsterm ist. Der auf der p-Norm basierende Regularisierungsterm lautet

$$y(\boldsymbol{h}) = \|\boldsymbol{a}\|_{p_f} \,, \tag{4.35}$$

wobei $\boldsymbol{a} = [a_1, \ldots, a_{L_a}]^T$ ein Vektor der Länge L_a ist. Die Wahl des Vektors $\boldsymbol{a}$ entscheidet über die Art der Regularisierung:

$$a_k = \begin{cases} H(e^{j\omega_k}) \,, & \text{zur Optimierung des Entzerrers,} \\ G(e^{j\omega_k}) \,, & \text{zur Optimierung des Gesamtsystems,} \end{cases} \tag{4.36}$$

wobei $H(e^{j\omega_k})$ die diskrete Fourier-Transformation des Entzerrers $h(n)$ an den diskreten Frequenzen ω_k, $k = 1, \ldots, L_a$ ist, beziehungsweise $G(e^{j\omega_k})$ die diskrete Fourier-Transformation der globalen Impulsantwort $g(n) = c(n) * h(n)$ an den diskreten Frequenzen ω_k, $k = 1, \ldots, L_a$ ist.

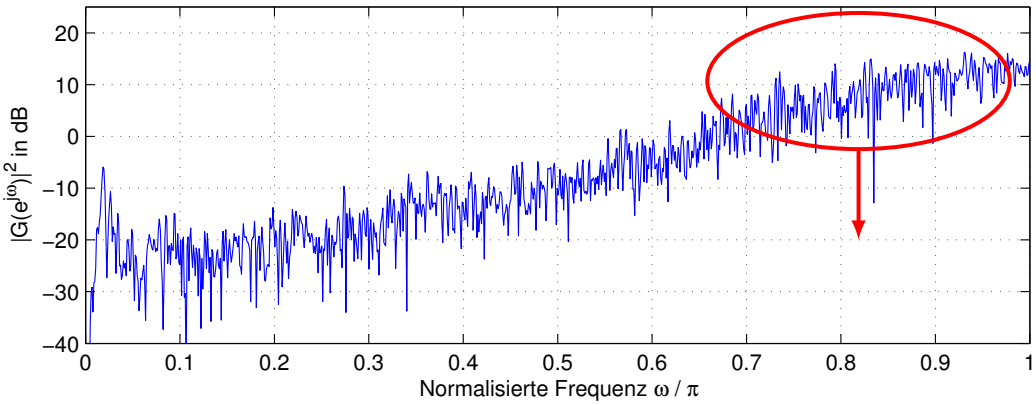

Abbildung 4.5: Schematische Darstellung der Auswirkung des auf der p-Norm basierenden Regularisierungsterms aus Gleichung (4.35): Starke Anhebungen im Betragsfrequenzgang der globalen Impulsantwort $g(n)$ werden bestraft.

Motivation Das in (4.35) formulierte Optimalitätskriterium bestraft Anhebungen im Betragsfrequenzgang. In Abbildung 4.5 ist schematisch die Bedeutung der Regularisierung dargestellt. Die Minimierung von (4.34) mit Werten $\alpha > 0$ führt somit zu einer Formung im Zeitbereich, während der Betragsfrequenzgang keine starken Anhebungen aufweist. Die Motivation den Frequenzgang des Entzerrer als Optimalitätskriterium zu betrachten ist die Überlegung, dass der Frequenzgang des Gesamtsystems keine starken Anhebungen aufweist, wenn der Frequenzgang des Entzerrers flach ist.

4.2.1.1 Herleitung des Gradienten

Das in (4.34) gegebene Optimierungsproblem wird durch ein Gradientenabstiegsverfahren gelöst. Die Lernregel bezüglich des Zielvektors $\boldsymbol{h}$ ist durch

$$\boldsymbol{h}^{l+1} = \boldsymbol{h}^{l} - \mu^{l}\left(\nabla_{\boldsymbol{h}} f\left(\boldsymbol{h}^{l}\right) + \alpha \nabla_{\boldsymbol{h}} y\left(\boldsymbol{h}^{l}\right)\right) \tag{4.37}$$

gegeben, wobei $\nabla_{\boldsymbol{h}} f(\boldsymbol{h})$ in Gleichung (3.66) gegeben ist. Der Gradient $\nabla_{\boldsymbol{h}} y(\boldsymbol{h})$ für den Regularisierungsterm (4.35) wird im Folgenden hergeleitet.

Gradient für die Optimierung der p-Norm Wird $p_f < \infty$ gewählt, so kann der in (4.35) formulierte Regularisierungsterm durch

$$y(\boldsymbol{h}) = \left(\sum_{k=1}^{L_a} |a_k|^{p_f}\right)^{\frac{1}{p_f}} \tag{4.38}$$

berechnet werden, wobei a_k die Koeffizienten der diskreten Fourier-Transformation des Entzerrers, beziehungsweise der globalen Impulsantwort enthält (siehe Gleichung (4.36)).

Die diskrete Fourier-Transformation kann als Matrix-Multiplikation beschrieben werden [119]. Mit $\boldsymbol{\mathcal{F}}$ als Matrix der diskreten Fourier-Transformation [119] wird der Vektor $\boldsymbol{a}$ als

$$\boldsymbol{a} = \boldsymbol{\mathcal{F}}\boldsymbol{T}\boldsymbol{h} \tag{4.39}$$

berechnet. Die Wahl von $\boldsymbol{T}$ entscheidet, ob der Betragsfrequenzgang des Entzerrers oder der Betragsfrequenzgang der globalen Impulsantwort optimiert werden soll. Es gilt

$$\boldsymbol{T} = \begin{cases} \boldsymbol{I}, & \text{zur Optimierung von } h(n), \\ \boldsymbol{C}, & \text{zur Optimierung von } g(n), \end{cases} \tag{4.40}$$

wobei $\boldsymbol{I}$ die Einheitsmatrix und $\boldsymbol{C}$ die Faltungsmatrix der Raumimpulsantwort $c(n)$ bezeichnet.

Der für das Optimierungsverfahren (4.37) benötigte Gradient des Regularisierungsterms (4.35) bezüglich des Zielvektors $\boldsymbol{h}$ ist durch

$$\nabla_{\boldsymbol{h}} y(\boldsymbol{h}) = \Re\left\{ \left(\sum_{k=1}^{L_a} |a_k|^{p_f} \right)^{\frac{1}{p_f}-1} (\boldsymbol{\mathcal{F}}\boldsymbol{T})^H \boldsymbol{b} \right\} \tag{4.41}$$

mit

$$\boldsymbol{b} = \operatorname{diag}\{\operatorname{sign}\{\boldsymbol{a}\}\} \, |\boldsymbol{a}|^{p_f-1} \tag{4.42}$$

gegeben [81].

Gradient für die Optimierung der Maximumsnorm Wird p_f in (4.35) als $p_f = \infty$ gewählt, so lässt sich (4.35) als

$$y(\boldsymbol{h}) = \|\boldsymbol{a}\|_\infty = \max\{|\boldsymbol{a}|\} \tag{4.43}$$

berechnen. Mit I_M als Index für den betragsmäßig größten Eintrag aus $\boldsymbol{a}$ gilt

$$y(\boldsymbol{h}) = \|\boldsymbol{a}\|_\infty = \max\{|\boldsymbol{a}|\} = |a_{I_M}| \,. \tag{4.44}$$

Zur Berechnung des Gradienten wird ein Vektor $\tilde{\boldsymbol{a}} = [\tilde{a}_1, \ldots, \tilde{a}_{L_a}]^T$ mit

$$\tilde{a}_k = \begin{cases} a_k, & \text{falls } k = I_M, \\ 0, & \text{sonst,} \end{cases} \tag{4.45}$$

definiert. Der Gradient von (4.35) für $p_f = \infty$ bezüglich des Zielvektors $\boldsymbol{h}$ ist dann durch

$$\nabla_{\boldsymbol{h}} y(\boldsymbol{h}) = \Re\{(\boldsymbol{\mathcal{F}T})^H \operatorname{sign}\{\tilde{\boldsymbol{a}}\}\} \tag{4.46}$$

gegeben [81].

4.2.1.2 Erweiterung der auf der p-Norm basierenden Regularisierung auf beliebig viele Kanäle

Der in Abschnitt 4.2.1 erarbeitete und in [81] publizierte Regularisierungsterm wurde für $p < \infty$ auf beliebig viele Übertragungskanäle erweitert [78, 79]. Dadurch ist eine Integration in das in Abschnitt 4.1 erarbeitete Formelwerk möglich.

Der Regularisierungsterm für ein Quellensignal $s_q(n)$ ist analog zu (4.35) als

$$y_q(\boldsymbol{h}) = \|\boldsymbol{a}_q\|_{p_f} \tag{4.47}$$

gegeben. Der Vektor $\boldsymbol{a}$ enthält für den Fall einer mehrkanaligen Übertragung die Konkatenation der diskreten Fourier-Transformierten aller $R \cdot M$ Realisierungen der globalen Impulsantworten $g_{mq}^{(r)}(n)$, $r = 1, \ldots, R$, $m = 1, \ldots, M$.

Das Optimierungsproblem mit dem auf der p-Norm basierenden Regularisierungsterm (4.47) lautet

$$\underset{\boldsymbol{h}_q}{\text{minimize}} \;:\; f_q(\boldsymbol{h}_q) + \alpha y_q(\boldsymbol{h}_q) \;\; \text{s.t.} \;\; \|\boldsymbol{h}_q\|_2 = 1, \tag{4.48}$$

wobei $f_q(\boldsymbol{h}_q)$ in (4.19) gegeben ist; $\alpha \geq 0$ ist der Gewichtsfaktor für den Regularisierungsterm $y_q(\boldsymbol{h}_q)$.

Das in (4.48) gegebene Optimierungsproblem wird durch ein Gradientenabstiegsverfahren gelöst. Die Lernregel lautet

$$\boldsymbol{h}_q^{l+1} = \boldsymbol{h}_q^l - \mu^l \left(\nabla_{\boldsymbol{h}_q} f_q\left(\boldsymbol{h}_q^l\right) + \alpha \nabla_{\boldsymbol{h}_q} y_q\left(\boldsymbol{h}_q^l\right) \right), \tag{4.49}$$

wobei l der Iterationsschritt und μ^l die adaptive Schrittweite im l-ten Iterationsschritt ist. Der Gradient $\nabla_{\boldsymbol{h}_q} f_q(\boldsymbol{h}_q)$ für die Formung im Zeitbereich ist in (4.25) gegeben, der Gradient $\nabla_{\boldsymbol{h}_q} y_q\left(\boldsymbol{h}_q^l\right)$ wird im Folgenden hergeleitet. Die Nebenbedingung wird durch eine

Normalisierung des Zielvektors $\boldsymbol{h}_q$ nach jeder Iteration des Gradientenabstiegsverfahrens hergestellt.

Herleitung des Gradienten Analog zu der in Abschnitt 4.1.1 verwendeten Notation wird eine Matrix $\widetilde{\boldsymbol{T}}$ als

$$\widetilde{\boldsymbol{T}} = \begin{cases} \boldsymbol{I}, & \text{zur Optimierung der Entzerrer} \\ \widetilde{\boldsymbol{C}}, & \text{zur Optimierung der Gesamtsysteme,} \end{cases} \tag{4.50}$$

mit

$$\widetilde{\boldsymbol{C}} = \begin{bmatrix} \boldsymbol{C}^{(1)} \\ \vdots \\ \boldsymbol{C}^{(R)} \end{bmatrix} \tag{4.51}$$

definiert, wobei $\boldsymbol{C}^{(r)}$ in (4.14) gegeben ist.

Des Weiteren wird eine Block-Diagonal-Matrix $\widetilde{\boldsymbol{\mathcal{F}}}$ der Form

$$\widetilde{\boldsymbol{\mathcal{F}}} = \operatorname{diag}\{\boldsymbol{\mathcal{F}}, \ldots, \boldsymbol{\mathcal{F}}\} \tag{4.52}$$

definiert, wobei $\boldsymbol{\mathcal{F}}$ die Matrix der diskreten Fourier-Transformation [119] kompatibler Größe ist. Mit $\widetilde{\boldsymbol{\mathcal{F}}}$ kann der Regularisierungsterm (4.47) als

$$y_q(\boldsymbol{h}_q) = \|\boldsymbol{a}_q\|_{p_f} = \left\|\widetilde{\boldsymbol{\mathcal{F}}}\widetilde{\boldsymbol{T}}\boldsymbol{h}_q\right\|_{p_f} \tag{4.53}$$

geschrieben werden. Die Berechnung des Gradienten von (4.47) ist damit analog zu (4.41) als

$$\nabla_{\boldsymbol{h}_q} y(\boldsymbol{h}) = \zeta_q(\boldsymbol{h}_q) \cdot \Re\left\{\left(\widetilde{\boldsymbol{\mathcal{F}}}\widetilde{\boldsymbol{T}}\right)^H \boldsymbol{b}_q\right\} \tag{4.54}$$

mit

$$\zeta_q(\boldsymbol{h}_q) = \left(\sum_{k=1}^{L_a} |a_{qk}|^{p_f}\right)^{\frac{1}{p_f}-1} \tag{4.55}$$

und

$$\boldsymbol{b}_q = \operatorname{diag}\{\operatorname{sign}\{\boldsymbol{a}_q\}\}\, |\boldsymbol{a}_q|^{p_f-1} \tag{4.56}$$

gegeben; in (4.55) bezeichnet a_{qk} das k-te Element des Vektors $\boldsymbol{a}_q$ aus (4.53).

4.2.2 Spektrale Flachheit als Kriterium

Ein Verfahren zur Optimierung des Betragsfrequenzgangs auf Basis des *spectral flatness measure* (SFM, siehe Abschnitt 2.3.2) wurde im Rahmen dieser Arbeit entwickelt und erstmalig in [74] publiziert. Für eine weitere Veröffentlichung wurde das SFM zu einem generalisierten Maß erweitert und als zusätzliches Optimalitätskriterium in den Filterentwurf integriert [75].

Zur besseren Übersicht wird an dieser Stelle noch einmal die Formel (2.21) zur Berechnung des *spectral flatness measures* gegeben:

$$\mathrm{SFM}_{g(n)} = \frac{\left(\prod\limits_{k=1}^{K} \left|G(e^{j\omega_k})\right|^2\right)^{\frac{1}{K}}}{\frac{1}{K}\sum\limits_{k=1}^{K} \left|G(e^{j\omega_k})\right|^2}, \tag{4.57}$$

wobei $G(e^{j\omega_k})$ die diskrete Fourier-Transformierte der Impulsantwort $g(n)$ an den K diskreten Frequenzen ω_k, $k = 1, \ldots, K$, ist.

Spektrale Flachheit als Kriterium für den Filterentwurf Im Falle eines *absolut flachen* Betragsfrequenzgangs

$$\left|G\left(e^{j\omega_1}\right)\right| = \ldots = \left|G\left(e^{j\omega_K}\right)\right|, \tag{4.58}$$

ist die Äquivalenz von geometrischem und arithmetischem Mittel (also Zähler und Nenner) in (4.57) gegeben; in diesem Fall ist das SFM gleich eins. Mit zunehmender Verzerrung des Betragsfrequenzgangs strebt das SFM gegen null.

Dieser Zusammenhang motiviert die Verwendung des SFMs als zusätzliches Optimalitätskriterium für den Filterentwurf, um ein Gesamtsystems mit einem flachen Betragsfrequenzgang zu erhalten.

4.2.2.1 Generalisiertes Maß zur Bewertung der spektralen Flachheit

Das *spectral flatness measure* (4.57) wurde im Rahmen dieser Arbeit durch die Integration zweier zusätzlicher Konzepte generalisiert. Zum einen kann das quadrierte Betragsspektrum vor der Bewertung in S einzelne Teilbänder integriert werden. Die Berechnung des geometrischen und des arithmetischen Mittels erfolgt dann nicht mehr über alle K einzelne Spektralkoeffizienten, sondern nur noch über die Energie in den S Teilbändern. Zum anderen kann jedes Teilband individuell gewichtet werden. Das

generalisierte Maß zur Bewertung der spektralen Flachheit wird als *generalized spectral flatness measure* (gSFM) bezeichnet und berechnet sich als

$$\mathrm{gSFM}_{g(n)} = \frac{\left(\prod_{s=1}^{S}\left(\gamma_s \sum_{k=1}^{K} w_{sk} \left|G(e^{j\omega_k})\right|^2\right)\right)^{\frac{1}{S}}}{\frac{1}{S}\sum_{s=1}^{S}\gamma_s \sum_{k=1}^{K} w_{sk} \left|G(e^{j\omega_k})\right|^2}, \tag{4.59}$$

wobei K die Anzahl der Spektralkoeffizienten $G(e^{j\omega_k})$, $k = 1, \ldots, K$, der Übertragungsfunktion ist. Die K Spektralkoeffizienten werden durch die Koeffizienten $w_{sk} \geq 0$, $s = 1, \ldots, S$, in S Teilbänder integriert und jedes Teilband wird vor der Berechnung des geometrischen (beziehungsweise des arithmetischen) Mittels mit einem Faktor $\gamma_s > 0$, $s = 1, \ldots, S$, individuell gewichtet.

Wird $S = K$, $\gamma_s = 1$, $s = 1, \ldots, S$, und $w_{sk} = \delta_{sk}$ als Kronecker-Delta (1.7) gewählt, so ist das gSFM (4.59) äquivalent zu dem SFM (4.57).

Das gSFM wurde im Verlauf dieser Arbeit entwickelt und erstmalig in [75] publiziert (dort als *integrated spectral flatness measure* bezeichnet).

Eigenschaften des generalisierten SFMs Das SFM nach (4.57) ist für den Fall maximal, dass der Betrag der Übertragungsfunktion *flach* ist (4.58) [105]. Im Gegensatz zu dem auf der p-Norm basierenden Regularisierungsterm (4.35) aus Abschnitt 4.2.1 werden somit nicht nur Anhebungen, sondern auch Einbrüche im Betragsfrequenzgang vermieden.

Die Energie in den Teilbändern sowie die individuelle Gewichtung der Teilbänder in (4.59) führt zu einer veränderten Bedingung, damit das gSFM maximal wird. Das gSFM ist genau dann maximal, wenn Zähler und Nenner, also geometrisches und arithmetisches Mittel der gewichteten Energie der S Teilbänder, gleich sind. Es gilt

$$\mathrm{gSFM}_{g(n)} = 1 \Longleftrightarrow \gamma_s \sum_{k=1}^{K} w_{sk} \left|G\left(e^{j\omega_k}\right)\right|^2 = v, \quad \forall s = 1, \ldots, S, \tag{4.60}$$

wobei v ein beliebiges Skalar größer null ist.

Gleichung (4.60) führt zu einer weiteren Überlegung. Damit (4.60) näherungsweise erfüllt ist, muss für die S Teilbänder gelten

$$\sum_{k=1}^{K} w_{sk} \left|G\left(e^{j\omega_k}\right)\right|^2 \approx \nu\gamma_s^{-1}, \quad s = 1, \ldots, S, \quad \nu > 0. \tag{4.61}$$

Durch die Wahl der Gewichtsfaktoren γ_s, $s = 1, \ldots, S$, kann also direkt die Gewichtung der Teilbänder zueinander bestimmt werden. Wird $S = K$ sowie $w_{sk} = \delta_{sk}$ nach (1.7) gewählt, so kann durch die Wahl der Gewichtsfaktoren γ_s direkt ein gewünschter Betragsfrequenzgang vorgegeben werden [75].

4.2.2.2 Regularisierung durch das generalisierte SFM

An dieser Stelle wird das gSFM als zusätzliches Optimalitätskriterium in das auf der p-Norm basierende Optimierungsproblem (3.57) aus [120] integriert. Der Regularisierungsterm lautet

$$s(\boldsymbol{h}) = -\log\big(\mathrm{gSFM}_{g(n)}\big)\,, \tag{4.62}$$

wobei der Logarithmus eine kompakte Ableitung des benötigten Gradienten ermöglicht [74, 75].

Das sich ergebende Optimierungsproblem lautet

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad : \quad f(\boldsymbol{h}) + \alpha s(\boldsymbol{h})\,, \tag{4.63}$$

wobei $\alpha \geq 0$ ein Gewichtsfaktor für den Regularisierungsterm (4.62) ist und $f(\boldsymbol{h})$ in (3.57) gegeben ist.

Die Optimierung von (4.63) erfolgt wie in [120] durch ein Gradientenabstiegsverfahren. Die Adaptationsregel lautet

$$\boldsymbol{h}^{l+1} = \boldsymbol{h}^{l} - \mu^{l}\big(\nabla_{\boldsymbol{h}} f\big(\boldsymbol{h}^{l}\big) + \alpha \nabla_{\boldsymbol{h}} s\big(\boldsymbol{h}^{l}\big)\big)\,, \tag{4.64}$$

wobei $\nabla_{\boldsymbol{h}} f(\boldsymbol{h})$ in (3.66) gegeben ist und μ^l die adaptive Schrittweite des Gradientenabstiegsverfahrens in Iteration l bezeichnet. Der benötigte Gradient für den Regularisierungsterm wird im Folgenden hergeleitet.

4.2.2.3 Gradient für das generalisierte SFM

Durch Anwendung der Logarithmengesetze kann (4.62) als

$$s(\boldsymbol{h}) = -\log\big(\mathrm{gSFM}_{g(n)}\big) = B(\boldsymbol{h}) - A(\boldsymbol{h}) \tag{4.65}$$

geschrieben werden, wobei

$$A(\boldsymbol{h}) = \log\left(\frac{1}{S}\sum_{s=1}^{S}\gamma_s\sum_{k=1}^{K}w_{sk}\left|G\left(e^{j\omega_k}\right)\right|^2\right) \tag{4.66}$$

die Berechnung des arithmetischen und

$$B(\boldsymbol{h}) = \log\left(\left(\prod_{s=1}^{S}\left(\gamma_s\sum_{k=1}^{K}w_{sk}\left|G\left(e^{j\omega_k}\right)\right|^2\right)\right)^{\frac{1}{S}}\right) \tag{4.67}$$

die Berechnung des geometrischen Mittels enthält. Die Berechnung des Gradienten $\nabla_{\boldsymbol{h}}s(\boldsymbol{h}) = \nabla_{\boldsymbol{h}}B(\boldsymbol{h}) - \nabla_{\boldsymbol{h}}A(\boldsymbol{h})$ erfolgt durch die getrennte Ableitung von (4.66) und (4.67) bezüglich des Zielvektors $\boldsymbol{h}$.

Für die Berechnung der Gradienten $\nabla_{\boldsymbol{h}}A(\boldsymbol{h})$ und $\nabla_{\boldsymbol{h}}B(\boldsymbol{h})$ wird Matrix-Vektor-Notation verwendet. Hierfür wird ein Vektor $\boldsymbol{g} = [g_1, \ldots, g_K]^T$ definiert, der die Einträge $g_k = G(e^{j\omega_k})$, $k = 1, \ldots, K$, für die diskreten Frequenzen ω_k enthält.

Es gilt

$$\boldsymbol{g} = \boldsymbol{M}\boldsymbol{h} \quad \text{mit} \quad \boldsymbol{M} = \widetilde{\boldsymbol{\mathcal{F}}}\boldsymbol{C}, \tag{4.68}$$

wobei $\widetilde{\boldsymbol{\mathcal{F}}} \in \mathbb{C}^{K\times L_g}$ die diskrete Fourier-Transformation für die K betrachteten Frequenzen ω_k beschreibt und $\boldsymbol{C}$ die Faltungsmatrix der Raumimpulsantwort $c(n)$ ist.

Die Integration der quadrierten Beträge der Übertragungsfunktion in Teilbänder sowie deren individuelle Gewichtung kann durch zwei weitere Matrizen ausgedrückt werden. Hierfür enthalte der Vektor $\widetilde{\boldsymbol{g}} = [\widetilde{g}_1, \ldots, \widetilde{g}_K]^T$ die quadrierten Beträge der Übertragungsfunktion, also $\widetilde{g}_k = |g_k|^2 = |G(e^{j\omega_k})|^2$. Wird eine Matrix $\boldsymbol{W}$ der Dimension $S \times K$ als

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & \cdots & w_{1K} \\ \vdots & \ddots & \vdots \\ w_{S1} & \cdots & w_{SK} \end{bmatrix} \tag{4.69}$$

sowie eine Diagonalmatrix $\boldsymbol{\Gamma}$ der Dimension $S \times S$ als

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_S \end{bmatrix} \tag{4.70}$$

definiert, so lässt sich das gSFM als

$$\mathrm{gSFM}_{g(n)} = \frac{\mathrm{geomean}(\boldsymbol{\Gamma W}\widetilde{\boldsymbol{g}})}{\mathrm{arithmean}(\boldsymbol{\Gamma W}\widetilde{\boldsymbol{g}})} \tag{4.71}$$

berechnen; $\mathrm{geomean}(\cdot)$ berechnet hierbei das geometrische Mittel eines Vektors und $\mathrm{arithmean}(\cdot)$ berechnet das arithmetisch Mittel eines Vektors.

Ausgehend von (4.71) kann die Integration und Gewichtung der quadrierten Beträge der Übertragungsfunktion in $\widetilde{\boldsymbol{g}}$ durch eine Matrix

$$\boldsymbol{\Psi} = \boldsymbol{\Gamma W} \tag{4.72}$$

der Dimension $S \times K$ mit den Einträgen ψ_{sk} beschrieben werden, wobei $\boldsymbol{\Gamma}$ und $\boldsymbol{W}$ in (4.70) und (4.69) gegeben sind.

Gradient für das arithmetische Mittel Ausgehend von (4.71), (4.65) und (4.72) kann (4.66) als

$$A(\boldsymbol{h}) = \log\left(\frac{1}{S}\sum_{s=1}^{S}\sum_{k=1}^{K}\psi_{sk}\left|g_k\right|^2\right) \tag{4.73}$$

geschrieben werden.

Durch Anwendung der Logarithmengesetze sowie der Kettenregel wird die partielle Ableitung von $A(\boldsymbol{h})$ bezüglich eines Filterkoeffizienten h_n innerhalb des Zielvektors $\boldsymbol{h}$ gebildet. Die partielle Ableitung lautet

$$\frac{\partial A(\boldsymbol{h})}{\partial h_n} = \zeta_A\left(\sum_{k=1}^{K} g_k m_{kn} \sum_{s=1}^{S}\psi_{sk}\right), \tag{4.74}$$

wobei m_{kn} das Element in der k-ten Zeile und der n-ten Spalte aus $\boldsymbol{M}$ (4.68) bezeichnet und ζ_A sich als

$$\zeta_A = \frac{2}{\sum_{s=1}^{S}\sum_{k=1}^{K}\psi_{sk}\left|g_k\right|^2} \tag{4.75}$$

berechnet.

Wird ein Vektor $\widetilde{\boldsymbol{\psi}} = \left[\widetilde{\psi}_1, \ldots, \widetilde{\psi}_K\right]^T$ mit den Einträgen

$$\widetilde{\psi}_k = \sum_{s=1}^{S}\psi_{sk} \tag{4.76}$$

definiert, so kann der Gradient letztendlich als

$$\nabla_{\boldsymbol{h}} A(\boldsymbol{h}) = \zeta_A \boldsymbol{M}^H \mathrm{diag}\left\{\widetilde{\boldsymbol{\psi}}\right\} \boldsymbol{g} \tag{4.77}$$

berechnet werden.

Gradient für das geometrische Mittel Mit (4.71), (4.65) und (4.72) kann (4.67) als

$$B(\boldsymbol{h}) = \log\left(\prod_{s=1}^{S}\left(\sum_{k=1}^{K}\psi_{sk}\left|g_k\right|^2\right)^{\frac{1}{S}}\right) \tag{4.78}$$

berechnet werden. Für die Berechnung des Gradienten $\nabla_{\boldsymbol{h}} B(\boldsymbol{h})$ werden wieder in einem ersten Schritt die partiellen Ableitungen von $B(\boldsymbol{h})$ bezüglich der einzelnen Filterkoeffizienten $h_n = h(n-1)$ gebildet. Die partielle Ableitung lautet

$$\frac{\partial B(\boldsymbol{h})}{\partial h_n} = \frac{2}{S}\sum_{s=1}^{S}\left(\zeta_{B_s}\sum_{k=1}^{K}\psi_{sk}\mathrm{sign}\{g_k\}\, m_{kn}\right), \tag{4.79}$$

mit

$$\zeta_{B_s} = \frac{1}{\sum_{k=1}^{K}\psi_{sk}\left|g_k\right|}. \tag{4.80}$$

Wird ein Vektor $\widehat{\boldsymbol{\psi}} = \left[\widehat{\psi}_1, \ldots, \widehat{\psi}_K\right]^T$ mit den Einträgen

$$\widehat{\psi}_k = \sum_{s=1}^{S}\zeta_{B_s}\psi_{sk} \tag{4.81}$$

definiert, so kann der Gradient $\nabla_{\boldsymbol{h}} B(\boldsymbol{h})$ letztendlich als

$$\nabla_{\boldsymbol{h}} B(\boldsymbol{h}) = \frac{2}{S}\boldsymbol{M}^H \mathrm{diag}\left\{\widehat{\boldsymbol{\psi}}\right\} \mathrm{sign}\{\boldsymbol{g}\} \tag{4.82}$$

berechnet werden.

4.2.3 Zusammenfassung

In diesem Abschnitt wurden zwei Verfahren erarbeitet die es ermöglichen, einen flachen beziehungsweise einen beliebig geformten Betragsfrequenzgang des Gesamtsystem zu fordern. Das in Abschnitt 4.2.1 entwickelte Verfahren basiert auf der p-Norm und

wurde in [81] publiziert. Dieses Verfahren wurde im weiteren Verlauf dieser Arbeit auf beliebig viele Übertragungskanäle und Abhörpositionen erweitert und in [78] sowie [79] veröffentlicht.

In Abschnitt 4.2.2 wurde ein weiteres Verfahren zur Regularisierung im Frequenzbereich erarbeitet. Die Methode basiert auf einer generalisierten Form des *spectral flatness measure* [105]. Das *generalized spectral flatness measure* wurde im Verlauf dieser Arbeit durch die Einführung des Konzeptes von Teilbändern mit beliebiger Gewichtung aus dem SFM entwickelt. Das erarbeitete Maß wurde in [75] publiziert. In dieser Arbeit wurde erstmalig ein Regularisierungsterm direkt auf Basis des SFM eingeführt, welcher sich durch die entsprechende Parametrierung aus dem generalisierten Maß ableiten lässt [74]; der auf dem gSFM basierende Regularisierungsterm wurde in [75] publiziert.

Für beide Verfahren wurden die benötigten Gradienten hergeleitet. Durch die speziellen Strukturen der beteiligten Matrizen können, im Fall von äquidistant verteilten diskreten Frequenzen ω_k, $k = 1, \ldots, K$, effiziente Verfahren wie die schnelle Fourier-Transformtion (FFT) und ihr Inverses (IFFT), sowie das Hadamard-Produkt für eine effiziente Implementierung verwendet werden.

4.3 Räumliche Robustheit

In ihren Arbeiten empfehlen Radlović et al., im Falle einer Fehlpositionierung von einer Entzerrung abzusehen, da diese schon bei kleinsten räumlichen Abweichungen üblicherweise zu einer massiven Verschlechterung der akustischen Eigenschaften des Gesamtsystems führt [139, 140]. In Abbildung 4.6 (oben) ist eine gemessene Raumimpulsantwort dargestellt. Die nach [81] geformte Impulsantwort ist in Abbildung 4.6 (mitte) dargestellt. Wird der für die Referenzposition entworfene Entzerrer auf eine Raumimpulsantwort, die wenige Millimeter neben der Referenzposition gemessen wurde, angewendet, so ergibt sich das in Abbildung 4.6 (unten) dargestellte Gesamtsystem. Es ist deutlich zu erkennen, dass keinerlei Enthallung mehr gegeben ist.

In diesem Abschnitt werden die neuen Methoden zur Erhöhung der Robustheit gegenüber räumlicher Fehlplatzierung entwickelt. Das Ziel ist, trotz einer Fehlplatzierung die Enthallung aufrecht zu erhalten.

In [114] wurden räumlich robuste Entzerrer auf Grundlage multipler Messungen der Raumimpulsantworten innerhalb des Abhörvolumens (siehe Abbildung 4.7.a) entworfen. Die Abtastintervalle werden hierbei durch das Abtasttheorem für Raumimpulsantworten [1, 114] bestimmt. Die praktische Anwendbarkeit dieser Methode scheitert an der

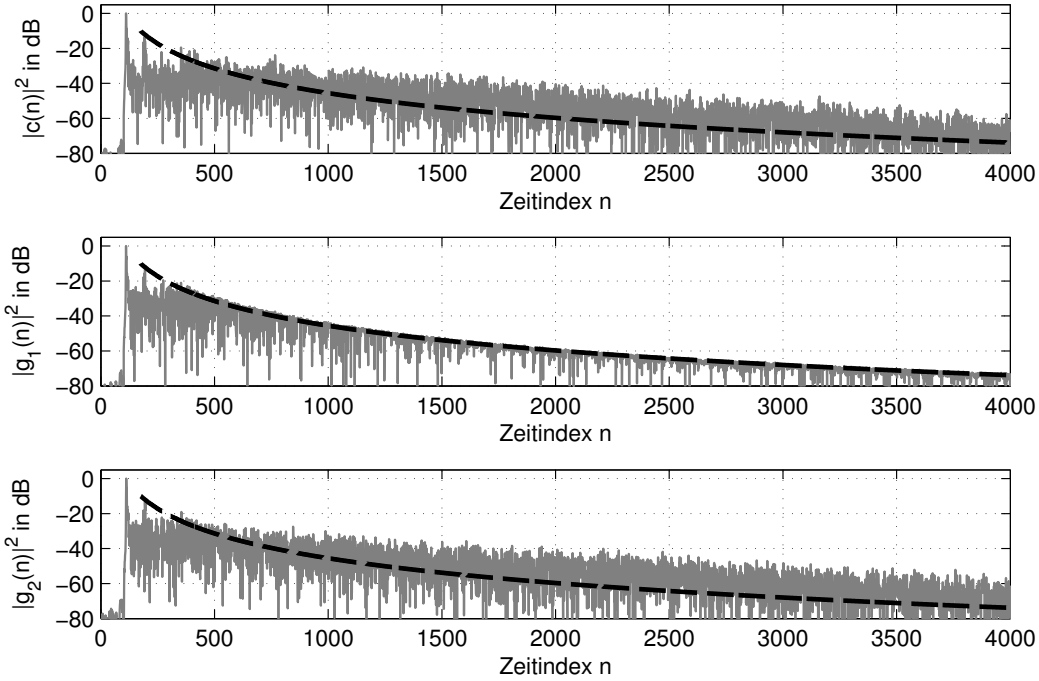

Abbildung 4.6: (Oben) An der Referenzposition gemessene Raumimpulsantwort. (Mitte) Geformte Impulsantwort an der Referenzposition. (Unten) Der für die Referenzposition entworfene Entzerrer wurde auf eine Raumimpulsantwort wenige Millimeter neben der Referenzposition angewendet.

aufwendigen Messung der Raumimpulsantworten. Im Rahmen dieser Arbeit wurden Verfahren entwickelt, die es erlauben, auf Grundlage einer einzelnen gemessenen Raumimpulsantwort räumlich robuste Entzerrer zu entwerfen. Der räumlich robuste Entwurf ermöglicht die Entzerrung innerhalb eines Raumvolumens auf Grundlage einer einzelnen Messung (siehe Abbildung 4.7.b).

4.3.1 Robustheit durch Fehlermodellierung

Im Rahmen dieser Arbeit wurde ein Verfahren zur Erhöhung der Robustheit gegenüber räumlicher Fehlplatzierung erarbeitet und in [79] sowie [76] publiziert. Hierfür werden die durch eine Fehlplatzierung induzierten Perturbationen der akustischen Kanäle statistisch beschrieben und während des Filterentwurfs berücksichtigt. Analog zu dem von Kallinger et al. in [82] vorgestellten Verfahren basiert die Beschreibung der Perturbationen auf den von Radlović et al. in [139, 140] formulierten statistischen Eigenschaften der Störungen. Ausgehend von einer einzelnen gemessenen Raumimpulsantwort pro Über-

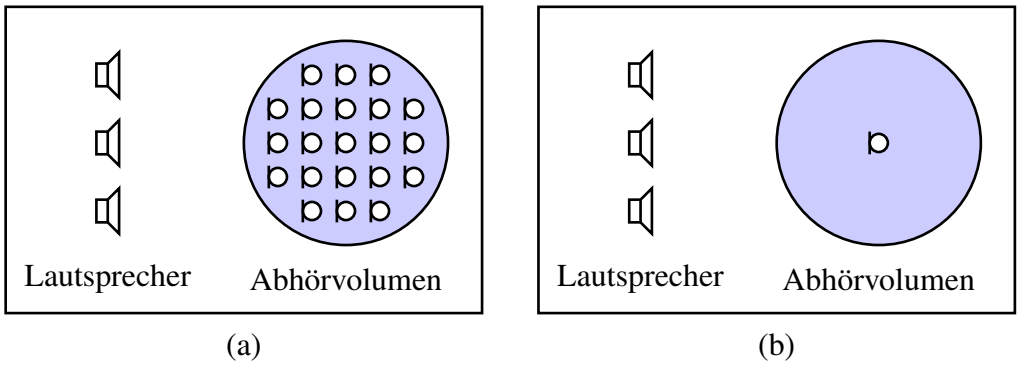

Abbildung 4.7: (a) Entwurf von räumlich robusten Entzerrern durch Abtastung des Abhörvolumens. (b) Entwurf räumlich robuster Entzerrer auf Grundlage einer einzelnen gemessenen Raumimpulsantwort, ohne den räumlichen Gültigkeitsbereich (hellblau) zu verkleinern.

tragungskanal, also von einem Lautsprecher zu dem Mikrofon an der Referenzposition, wird ein Fehlerterm entwickelt, der eine räumlich robuste Entzerrung ermöglicht.

4.3.1.1 Fehlermodell nach Radlović et al.

In ihren Arbeiten untersuchten Radlović et al. [139, 140] die frequenzabhängige Fehlerenergie im Falle einer Fehlplatzierung (siehe Abschnitt 3.4.4). Hierfür nehmen Radlović et al. eine *perfekte* Entzerrung an der Referenzposition an. Mit der Kreisfrequenz $\omega = 2\pi f$ lässt sich der frequenzabhängige Fehlerterm (3.85) als

$$U(\omega) = \mathrm{E}\big\{|(C(\omega) + P(\omega))\, H(\omega) - 1|^2\big\} \tag{4.83}$$

schreiben, wobei $C(\omega)$, $P(\omega)$ und $H(\omega)$ die Fourier-Transformierten der Raumimpulsantwort $c(t)$, der Perturbation $p(t)$ und des Entzerrers $h(t)$ bezeichnen.

Ausgehend von einer *perfekten* Entzerrung an der Referenzposition, also $H(\omega) = 1/C(\omega)$, und Fernfeldbedingungen ergibt sich für den frequenzabhängigen Fehlerterm (4.83) der Ausdruck

$$U(\omega) = \frac{\mathrm{E}\big\{|P(\omega)|^2\big\}}{|C(\omega)|^2} = 2 - 2\frac{\sin(\omega D/v)}{\omega D/v}, \tag{4.84}$$

mit v als Schallgeschwindigkeit und D als Distanz zur Referenzposition. Durch Umstellen von (4.84) ergibt sich

$$\mathrm{E}\{|P(\omega)|^2\} = |C(\omega)|^2 \left(2 - 2\frac{\sin(\omega D/v)}{\omega D/v}\right). \tag{4.85}$$

4.3.1.2 Herleitung des Fehlerterms

Ausgehend von bandbegrenzten Signalen mit einer Maximalfrequenz ω_c und unter Einhaltung des Abtasttheorems [149] (also Abtastrate $f_s = 1/T \geq \omega_c/\pi$) können die zeitkontinuierlichen Komponenten $c(t)$, $p(t)$ und $h(t)$ aus (4.83) durch ihre zeitdiskreten Äquivalente $c(n)$, $p(n)$ und $h(n)$ ersetzt werden.

Ausgehend von (4.85) kann die Autokorrelationsfolge der Perturbation $p(n)$ als

$$r_{pp}(n) = r_{cc}(n) * u(n) \tag{4.86}$$

mit $r_{cc}(n) = c(n) * c(-n)$ berechnet werden, wobei

$$u(n) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} \left(2 - 2\frac{\sin(\omega D/v)}{\omega D/v}\right) e^{j\omega nT}\, d\omega. \tag{4.87}$$

Die Folge $u(n)$ kann berechnet werden, indem $U(\omega)$ aus Gleichung (4.84) äquidistant an den diskreten Frequenzen $\omega_k = \frac{k\omega_c}{K}$, mit $k = -K, -K+1, \ldots, K$, abgetastet wird und danach durch die inverse diskrete Fourier-Transformation (IDFT) in den Zeitbereich überführt wird [79].

Mit (4.1) und (3.70), sowie bei der Verwendung von N Lautsprechern lässt sich die globale Impulsantwort als

$$g(n) = \sum_{\ell=1}^{N} [c_\ell(n) + p_\ell(n)] * h_\ell(n) \tag{4.88}$$

beschreiben, wobei $p_\ell(n)$ der durch Fehlplatzierung induzierte Perturbationsterm der Raumimpulsantwort von Lautsprecher ℓ, $\ell = 1, \ldots, N$, zu dem Mikrofon an der Referenzposition ist.

Wie in Abschnitt 3.2.1.3 beschrieben wird für die Formung die globale Impulsantwort $g(n)$ mit einem Gewichtsfenster $w(n)$ bewertet. Unter der Annahme, dass der akustische

Kanal an der Referenzposition vollständig entzerrt wurde [139] lässt sich der durch eine Fehlplatzierung verursachte Fehler als

$$e(n) = w(n) \left[\sum_{\ell=1}^{N} p_\ell(n) * h_\ell(n) \right] \tag{4.89}$$

berechnen [76]. Im Rahmen dieser Arbeit wird lediglich der ungewünschte Teil der globalen Impulsantwort betrachtet; für das Fenster $w(n)$ aus (4.89) gilt also $w(n) = w_u(n)$, wobei $w_u(n)$ wie in [120] empfohlen als Reziprokes der temporalen Maskierungsschwelle des menschlichen Hörapparates [35] gewählt wurde.

Mit $\boldsymbol{W} = \mathrm{diag}\{\boldsymbol{w}\}$, $\boldsymbol{w} = [w(0)\,, \ldots, w(L_g - 1)]^T$, $\boldsymbol{P}_\ell$ als Faltungsmatrix der Perturbation $p_\ell(n)$ und $\boldsymbol{h}_\ell = [h_\ell(0)\,, \ldots, h_\ell(L_h - 1)]^T$ ist der durch die Fehlplatzierung verursachte mittlere quadratische Fehler durch

$$O = \mathrm{E}\left\{ \sum_{\ell=1}^{N} \|\boldsymbol{W}\boldsymbol{P}_\ell \boldsymbol{h}_\ell\|_2^2 \right\} \tag{4.90}$$

gegeben. Für einen robusten Filterentwurf müssen die Filterkoeffizienten derart entworfen werden, dass (4.90) minimiert wird.

Berechnung des Fehlerterms Unter der Annahme, dass die Perturbationen jeweils paarweise unkorreliert sind und mit $\boldsymbol{H}_\ell$ als Faltungsmatrix des Filters $h_\ell(n)$ kann (4.90) als

$$O = \sum_{\ell=1}^{N} O_\ell \tag{4.91}$$

mit

$$\begin{aligned} O_\ell &= \mathrm{E}\big\{ \|\boldsymbol{W}\boldsymbol{P}_\ell \boldsymbol{h}_\ell\|_2^2 \big\} \\ &= \mathrm{E}\big\{ \boldsymbol{h}_\ell^T \boldsymbol{P}_\ell^T \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{P}_\ell \boldsymbol{h}_\ell \big\} \\ &= \mathrm{E}\big\{ \boldsymbol{p}_\ell^T \boldsymbol{H}_\ell^T \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{H}_\ell \boldsymbol{p}_\ell \big\} \end{aligned} \tag{4.92}$$

geschrieben werden [76, 79].

Zur weiteren Umformung von (4.92) werden die Eigenschaften der Spur einer Matrix ausgenutzt. Für die Spur (englisch *trace*, $\mathrm{tr}(\cdot)$) einer Matrix $\boldsymbol{A}$ gilt

$$\mathrm{E}\{\mathrm{tr}(\boldsymbol{A})\} = \mathrm{tr}(\mathrm{E}\{\boldsymbol{A}\})\,, \tag{4.93}$$

sowie die Invarianz gegenüber zirkulärer Permutation

$$\mathrm{tr}(\boldsymbol{ABC}) = \mathrm{tr}(\boldsymbol{BCA}) = \mathrm{tr}(\boldsymbol{CAB})\,, \tag{4.94}$$

wobei $\boldsymbol{A}$, $\boldsymbol{B}$ und $\boldsymbol{C}$ quadratische Matrizen sind [17].

Die Eigenschaften (4.93) und (4.94) erlauben, (4.92) neu zu formulieren:

$$\begin{aligned} O_\ell &= \mathrm{E}\big\{\mathrm{tr}\big(\boldsymbol{H}_\ell^T \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{H}_\ell \boldsymbol{p}_\ell \boldsymbol{p}_\ell^T\big)\big\} \\ &= \mathrm{tr}\Big(\boldsymbol{H}_\ell^T \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{H}_\ell \boldsymbol{R}_{pp}^{(\ell)}\Big)\,, \end{aligned} \tag{4.95}$$

wobei $\boldsymbol{R}_{pp}^{(\ell)} = \mathrm{E}\big\{\boldsymbol{p}_\ell \boldsymbol{p}_\ell^T\big\}$ die Autokorrelationsmatrix des Perturbationsterms $p_\ell(n)$ ist.

Für eine gegebene Raumimpulsantwort $c_\ell(n)$ und eine angenommene Fehlplatzierung D kann $\boldsymbol{R}_{pp}^{(\ell)}$ als Toeplitz-Matrix aus der Autokorrelationssequenz $r_{pp}^{(\ell)}(n)$ aus (4.86) aufgestellt werden.

Zur weiteren Umformung von (4.95) wird die Autokorrelationsmatrix $\boldsymbol{R}_{pp}^{(\ell)}$ mit Hilfe der Cholesky-Dekomposition [49, 117] derart faktorisiert, dass gilt

$$\boldsymbol{R}_{pp}^{(\ell)} = \boldsymbol{L}^{(\ell)} \boldsymbol{L}^{(\ell)^T}. \tag{4.96}$$

Mit der in Gleichung (4.96) gegebenen Faktorisierung der Autokorrelationsmatrix kann (4.95) als

$$O_\ell = \mathrm{tr}\Big(\boldsymbol{L}^{(\ell)^T} \boldsymbol{H}_\ell^T \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{H}_\ell \boldsymbol{L}^{(\ell)}\Big) \tag{4.97}$$

berechnet werden.

Bezeichne $\boldsymbol{l}_i^{(\ell)}$ die i-te Spalte der Matrix $\boldsymbol{L}^{(\ell)}$ und $\boldsymbol{L}_i^{(\ell)}$ die aus $\boldsymbol{l}_i^{(\ell)}$ generierte Faltungsmatrix. Mit $\boldsymbol{H}_\ell \boldsymbol{l}_i^{(\ell)} = \boldsymbol{L}_i^{(\ell)} \boldsymbol{h}_\ell$ und

$$\boldsymbol{M}_\ell = \sum_{i=1}^{L_c} \boldsymbol{L}_i^{(\ell)^T} \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{L}_i^{(\ell)} \tag{4.98}$$

kann (4.95) letztendlich als

$$O_\ell = \boldsymbol{h}_\ell^T \boldsymbol{M}_\ell \boldsymbol{h}_\ell \tag{4.99}$$

berechnet werden. Der in (4.99) gegebene Ausdruck kann bezüglich der gesuchten Filterkoeffizienten differenziert und somit einfach in den Filterentwurf integriert werden [76, 79].

4.3.1.3 Integration des Fehlerterms in den Filterentwurf

In diesem Abschnitt wird der in Abschnitt 4.3.1.2 hergeleitete und auf den Ergebnissen von Radlović et al. [139, 140] basierende Fehlerterm in das Formelwerk zum Filterentwurf integriert. Der stochastische Fehlerterm wird lediglich für den *ungewünschten* Teil der globalen Impulsantworten betrachtet. Hierbei wird, wie auch schon in Abschnitt 4.3.1.2, von der Reproduktion über mehrere Lautsprecher ausgegangen.

Die erweiterte Zielfunktion lautet

$$f(\boldsymbol{h}) = \log\left(\frac{\bar{f}_u(\boldsymbol{h})}{f_d(\boldsymbol{h})}\right), \tag{4.100}$$

wobei $f_d = \|\boldsymbol{g}_d\|_{p_d}$ mit $\boldsymbol{g}_d = \operatorname{diag}\{\boldsymbol{w}_d\} \sum\limits_{\ell=1}^{N} \boldsymbol{C}_\ell \boldsymbol{h}_\ell$ und $\bar{f}_u(\boldsymbol{h})$ durch die Integration des statistischen Fehlerterms aus Abschnitt 4.3.1.2 als

$$\bar{f}_u(\boldsymbol{h}) = \mathrm{E}\Big\{\|\boldsymbol{g}_u\|_{p_u}\Big\} \tag{4.101}$$

mit

$$\boldsymbol{g}_u = \operatorname{diag}\{\boldsymbol{w}_u\}([\boldsymbol{C}_1, \ldots, \boldsymbol{C}_N] + [\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N]) \begin{bmatrix} \boldsymbol{h}_1 \\ \vdots \\ \boldsymbol{h}_N \end{bmatrix} = \boldsymbol{W}_u(\boldsymbol{C} + \boldsymbol{P})\,\boldsymbol{h} \tag{4.102}$$

definiert wird [76, 79].

Für die p-Norm des Vektors $\boldsymbol{g}_u$ aus (4.102) besagt die Minkowski-Ungleichung [3], dass

$$\|\boldsymbol{W}_u\boldsymbol{C}\boldsymbol{h} + \boldsymbol{W}_u\boldsymbol{P}\boldsymbol{h}\|_{p_u} \leq \|\boldsymbol{W}_u\boldsymbol{C}\boldsymbol{h}\|_{p_u} + \|\boldsymbol{W}_u\boldsymbol{P}\boldsymbol{h}\|_{p_u}\,. \tag{4.103}$$

Die Minkowski-Ungleichung (4.103) motiviert, statt (4.100) lediglich die obere Grenze der Zielfunktion zu minimieren; hierfür wird $\bar{f}_u(\boldsymbol{h})$ aus (4.101) durch den Ausdruck

$$\|\boldsymbol{W}_u\boldsymbol{C}\boldsymbol{h}\|_{p_u} + \mathrm{E}\Big\{\|\boldsymbol{W}_u\boldsymbol{P}\boldsymbol{h}\|_{p_u}\Big\} \tag{4.104}$$

ersetzt [79].

Überführung in ein quadratisches Kriterium Da Radlović et al. [139, 140] keinerlei Aussagen über die statistischen Eigenschaften höherer Ordnung des Perturbati-

onsterms machen, wird im Rahmen dieser Arbeit lediglich ein quadratischer Fehlerterm in die Zielfunktion integriert. Die erarbeitete und in [79] publizierte Zielfunktion lautet

$$f(\boldsymbol{h}) = \log\left(\frac{\tilde{f}_u(\boldsymbol{h})}{f_d(\boldsymbol{h})}\right), \tag{4.105}$$

mit

$$\tilde{f}_u(\boldsymbol{h}) = \|\boldsymbol{W}_u\boldsymbol{C}\boldsymbol{h}\|_{p_u} + \beta\mathrm{E}\{\|\boldsymbol{W}_u\boldsymbol{P}\boldsymbol{h}\|_2\}, \tag{4.106}$$

wobei $\beta \geq 0$ als Gewichtsfaktor für den Regularisierungsterm verstanden werden kann [76, 79].

Optimierung der Zielfunktion Das Optimierungsproblem zum Entwurf von räumlich robusten Filtern lautet:

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad : \quad f(\boldsymbol{h}) + \alpha y(\boldsymbol{h}) \;\; \text{s.t.} \;\; \|\boldsymbol{h}\|_2 = 1, \tag{4.107}$$

wobei $f(\boldsymbol{h})$ in (4.105) gegeben ist und $y(\boldsymbol{h})$ in (4.53) für $Q = M = R = 1$ gegeben ist [79].

Zur Lösung des Optimierungsproblems wird ein Gradientenabstiegsverfahren angewendet, wobei die Nebenbedingung durch eine Normalisierung des Zielvektors nach jeder Iteration hergestellt wird. Der Gradient bezüglich der Filterkoeffizienten ist durch

$$\nabla_{\boldsymbol{h}} f(\boldsymbol{h}) = \frac{1}{\tilde{f}_u(\boldsymbol{h})}\nabla_{\boldsymbol{h}}\tilde{f}_u(\boldsymbol{h}) - \frac{1}{f_d(\boldsymbol{h})}\nabla_{\boldsymbol{h}} f_d(\boldsymbol{h}) \tag{4.108}$$

gegeben, wobei $\nabla_{\boldsymbol{h}} f_d(\boldsymbol{h})$ in (4.26) für $Q = M = R = 1$ gegeben ist und

$$\nabla_{\boldsymbol{h}}\tilde{f}_u(\boldsymbol{h}) = \nabla_{\boldsymbol{h}} f_u(\boldsymbol{h}) + \beta\nabla_{\boldsymbol{h}} f_u^{(P)}(\boldsymbol{h}) \tag{4.109}$$

mit $\nabla_{\boldsymbol{h}} f_u(\boldsymbol{h})$ in (4.30) für $Q = 1$, $M = 1$ und $R = 1$ gegeben ist. Der verbleibende Term des Gradienten, $\nabla_{\boldsymbol{h}} f_u^{(P)}(\boldsymbol{h})$, wird mit der in Abschnitt 4.3.1.2 erarbeiteten Formulierung berechnet.

Mit

$$f_u^{(P)}(\boldsymbol{h}) = \mathrm{E}\{\|\boldsymbol{W}_u\boldsymbol{P}\boldsymbol{h}\|_2\} = \left(\sum_{\ell=1}^{N} \boldsymbol{h}_\ell^T \boldsymbol{M}_\ell \boldsymbol{h}_\ell\right)^{\frac{1}{2}}, \tag{4.110}$$

wobei $\boldsymbol{M}_\ell$ in (4.98) gegeben ist, berechnet sich der benötigte Gradient als

$$\nabla_{\boldsymbol{h}} f_u^{(P)}(\boldsymbol{h}) = \frac{1}{2}\left(\sum_{\ell=1}^{N} \boldsymbol{h}_\ell^T \boldsymbol{M}_\ell \boldsymbol{h}_\ell\right)^{-\frac{1}{2}} \begin{bmatrix} \left(\boldsymbol{M}_1 + \boldsymbol{M}_1^T\right)\boldsymbol{h}_1 \\ \vdots \\ \left(\boldsymbol{M}_N + \boldsymbol{M}_N^T\right)\boldsymbol{h}_N \end{bmatrix}. \tag{4.111}$$

4.3.1.4 Erweiterung auf beliebig viele Kanäle

Das in diesem Abschnitt entwickelte Verfahren zum Entwurf robuster Filter wurde in [79] zur gleichzeitigen Enthallung und Übersprechkompensation verwendet. In diesem Abschnitt soll die Integration des Fehlerterms (4.90) in das Formelwerk zur mehrkanaligen Formung aus Abschnitt 4.1 präsentiert werden.

Ausgehend von einem allgemeingültigen Szenario mit Q Quellensignalen, M Abhörpositionen und N Lautsprechern werden die notwendigen Modifikationen vorgestellt; hierbei wird von lediglich einer Messung jedes akustischen Kanals ausgegangen, also $R = 1$.

Das hier entwickelte und in [79] veröffentlichte Optimierungsproblem zum Filterentwurf kann als Generalisierung von (4.107) auf beliebig viele Quellensignale und Abhörpositionen verstanden werden:

$$\underset{\boldsymbol{h}_q}{\text{minimize}} \quad : \quad \log\left(\frac{\tilde{f}_q^{(u)}(\boldsymbol{h}_q)}{f_q^{(d)}(\boldsymbol{h}_q)}\right) + \alpha y(\boldsymbol{h}_q) \quad \text{s.t.} \quad \|\boldsymbol{h}_q\|_2 = 1, \tag{4.112}$$

wobei $y(\boldsymbol{h}_q)$ der in Abschnitt 4.2.1 eingeführte und in Abschnitt 4.2.1.2 auf beliebig viele Übertragungskanäle erweiterte auf der p-Norm basierende Regularisierungsterm ist; $\alpha \geq 0$ ist der Gewichtsfaktor für den Regularisierungsterm. Der Term $f_q^{(d)}(\boldsymbol{h}_q)$ ist in (4.20) gegeben, und

$$\tilde{f}_q^{(u)}(\boldsymbol{h}_q) = f_q^{(u)}(\boldsymbol{h}_q) + \beta f_q^{(P)}(\boldsymbol{h}_q) \tag{4.113}$$

mit $f_q^{(u)}(\boldsymbol{h}_q)$ in (4.21) und

$$f_q^{(P)}(\boldsymbol{h}_q) = \mathrm{E}\left\{\left\|\widetilde{\boldsymbol{W}}_q \boldsymbol{P} \boldsymbol{h}_q\right\|_2\right\} = \left(\sum_{m=1}^{M}\sum_{\ell=1}^{N} \boldsymbol{h}_{\ell q}^T \boldsymbol{M}_{m\ell} \boldsymbol{h}_{\ell q}\right)^{\frac{1}{2}} \tag{4.114}$$

mit $\boldsymbol{P}$ in (4.12) und

$$\widetilde{\boldsymbol{W}}_q = \mathrm{diag}\left\{\left[\boldsymbol{w}_{1q}^T, \ldots, \boldsymbol{w}_{Mq}^T\right]\right\}, \tag{4.115}$$

wobei $\boldsymbol{w}_{mq}$ jeweils die Koeffizienten des Gewichtsfensters $w_{mq}^{(u)}(n)$ enthält.

Die Matrix $\boldsymbol{M}_{m\ell}$ aus (4.114) ist durch (4.98) gegeben, wobei die Gewichtsfenster und Raumimpulsantworten entsprechend für jedes Quellensignal $s_q(n)$ und Mikrofonposition m gewählt werden (siehe Abschnitt 4.4 sowie [79]).

Die Optimierung von (4.112) erfolgt durch ein Gradientenabstiegsverfahren. An dieser Stelle wird der Übersichtlichkeit halber lediglich der Gradient für die Erweiterung der Kostenfunktion angegeben.

Der Gradient für (4.113) lautet

$$\nabla_{\boldsymbol{h}_q} \tilde{f}_q^{(u)}(\boldsymbol{h}_q) = \nabla_{\boldsymbol{h}_q} f_q^{(u)}(\boldsymbol{h}_q) + \beta \nabla_{\boldsymbol{h}_q} f_q^{(P)}(\boldsymbol{h}_q)\,, \tag{4.116}$$

wobei $\nabla_{\boldsymbol{h}_q} f_q^{(u)}(\boldsymbol{h}_q)$ in (4.30) gegeben ist und

$$\nabla_{\boldsymbol{h}_q} f_q^{(P)}(\boldsymbol{h}_q) = \frac{1}{2}\left(\sum_{m=1}^{M}\sum_{\ell=1}^{N} \boldsymbol{h}_{\ell q}^T \boldsymbol{M}_{m\ell} \boldsymbol{h}_{\ell q}\right)^{-\frac{1}{2}} \begin{bmatrix} \sum_{m=1}^{M}\left(\boldsymbol{M}_{m1} + \boldsymbol{M}_{m1}^T\right)\boldsymbol{h}_{1q} \\ \vdots \\ \sum_{m=1}^{M}\left(\boldsymbol{M}_{mN} + \boldsymbol{M}_{mN}^T\right)\boldsymbol{h}_{Nq} \end{bmatrix}. \tag{4.117}$$

4.3.1.5 Effiziente Implementierung

Die Berechnung von (4.98) ist sehr zeitaufwendig. Der Algorithmus zur Berechnung der benötigten Matrizen aus [79] basiert auf der Cholesky-Dekomposition [49, 117] von Matrizen der Dimension $L_c \times L_c$ und einer anschließenden elementweisen Summation von L_c Matrizen der Dimension $L_h \times L_h$ [76].

Im Verlauf dieser Arbeit wurde ein effizientes Verfahren zur Berechnung der benötigten Matrizen $\boldsymbol{M}_\ell$ (beziehungsweise $\boldsymbol{M}_{q\ell}$) entwickelt, welches zu einer massiven Beschleunigung der Vorberechnungen führt und damit essentiell für eine praktische Nutzbarmachung des Verfahrens ist [76] .

Herleitung der Berechnung Werden die Strukturen der Matrizen $\boldsymbol{W}$, $\boldsymbol{H}_\ell$ und $\boldsymbol{R}_{pp}^{(\ell)}$ in (4.95) bei der Berechnung des Fehlerterms beachtet, so kann der Fehlerterm O_ℓ aus (4.95) als

$$O_\ell = \sum_{n=0}^{L_g-1}\sum_{i=0}^{L_c-1}\sum_{j=0}^{L_c-1} h_\ell(n-j)\, h_\ell(n-i)\, r_{pp}^{(\ell)}(i-j)\, w^2(n) \tag{4.118}$$

berechnet werden. Durch Gleichsetzen von (4.118) mit (4.99) wurde eine einfache Regel zur Berechnung der Matrix $\boldsymbol{M}_\ell$ aus (4.99) gefunden. Da $\boldsymbol{M}_\ell$ symmetrisch ist muss

lediglich die obere rechte Dreiecksmatrix berechnet werden und die entsprechenden Einträge sind in die untere linke Hälfte zu kopieren.

Effiziente Berechnungsvorschrift Bezeichne $[\boldsymbol{M}_\ell]_{p,q}$, $1 \leq p, q \leq L_h$ das Element in der p-ten Zeile und q-ten Spalte der Matrix $\boldsymbol{M}_\ell$. Das Element $[\boldsymbol{M}_\ell]_{p,q}$ ist durch

$$[\boldsymbol{M}_\ell]_{p,q} = \begin{cases} \sum\limits_{n=q-1}^{p+L_c-2} w^2(n)\, r_{pp}^{(\ell)}(q-p)\,, & q \geq p \\ [\boldsymbol{M}_\ell]_{q,p}\,, & \text{sonst} \end{cases} \tag{4.119}$$

gegeben [76].

4.3.2 Perturbationsbasierter Ansatz

In diesem Abschnitt wird ein weiteres Verfahren entwickelt, um eine räumlich robuste Entzerrung zu erreichen [73, 77]. Die Methode verbindet die statistische Beschreibung der Systemabweichungen aus [79] sowie [76] mit der Methode der multiplen Messungen von Raumimpulsantworten innerhalb eines kleinen Volumens aus [78, 114] sowie [79].

Das Verfahren basiert darauf, R zufällig perturbierte Instanzen der gemessenen Raumimpulsantwort zu generieren und diese als zusätzliche Realisierungen des akustischen Kanals in die Filteroptimierung zu integrieren. Das erarbeitete Modell beschreibt die Systemperturbationen als nicht-stationären Rauschterm, welcher sich aus dem Abklingmodell nach Polack [135] ergibt [73]; als Verfeinerung wurden die Ergebnisse von Radlović et al. [139, 140] in das Fehlermodell integriert. Das Verfahren wurde in [77] publiziert.

4.3.2.1 Beschreibung der Perturbationen

Zur Herleitung des erarbeiteten Modells wird an dieser Stelle von zeitkontinuierlichen Impulsantworten ausgegangen. Eine perturbierte Raumimpulsantwort $\hat{c}(t)$ wird in dem erarbeiteten Modell als

$$\hat{c}(t) = \rho c(t) + A p(t) \tag{4.120}$$

beschrieben; hierbei ist $c(t)$ die gemessene Raumimpulsantwort, $p(t)$ beschreibt den additiven Störterm. Die Faktoren $0 \leq \rho \leq 1$ und $0 \leq A$ ermöglichen es, die Stärke der Perturbation $p(t)$ festzulegen und gleichzeitig die plausible Forderung, dass eine Raum-

impulsantwort die gleiche Energie wie die perturbierte (beziehungsweise benachbarte) Raumimpulsantwort hat, zu erfüllen.

In vielen Veröffentlichungen werden die Perturbationen als stationäres weißes Rauschen angenommen (siehe Abschnitt 3.4). Diese Annahme führt im *least-squares* basierten Filterentwurf letztendlich zu der Forderung, dass die zu entwerfenden Filter eine möglichst geringe Energie haben sollen [63, 88].

Im Folgenden werden die Eigenschaften der Perturbationen im Zeit- sowie im Frequenzbereich hergeleitet. Diese Eigenschaften bilden die Grundlage für die Synthese künstlich perturbierter Realisierungen der Raumimpulsantwort zur Referenzposition.

4.3.2.2 Modellierung der Perturbation im Zeitbereich

Für die Herleitung der Beschreibung der Perturbation $p(t)$ im Zeitbereich wird das Abklingmodell nach Polack [135] verwendet. In seiner Arbeit beschreibt Polack eine Raumimpulsantwort als eine Realisierung eines nicht-stationären Rauschprozesses:

$$c(t) = \begin{cases} b(t)\, e^{-\Delta t}, & t \geq 0 \\ 0, & t < 0, \end{cases} \tag{4.121}$$

wobei $b(t)$ ein mittelwertfreies, stationäres, gaussförmiges Rauschsignal ist; der Einfachheit halber wird eine Varianz von eins ($\sigma^2_{b(t)} = 1$) angenommen. Der Wert Δ ist durch

$$\Delta = \frac{3\ln(10)}{T_{60}} \tag{4.122}$$

mit der Nachhallzeit T_{60} [98] des zu modellierenden Raumes verknüpft [135].

Mit (4.121) ist die Einhüllende einer quadrierten Raumimpulsantwort als

$$\mathrm{E}\{c^2(t)\} = e^{-2\Delta t} \tag{4.123}$$

gegeben.

Für das erarbeitete Modell wird angenommen, dass eine Raumimpulsantwort $c(t)$ sowie eine in der Nähe von $c(t)$ gemessene Raumimpulsantwort $\hat{c}(t)$ das gleiche Abklingverhalten aufweisen:

$$\mathrm{E}\{c^2(t)\} = \mathrm{E}\{\hat{c}^2(t)\} = e^{-2\Delta t}. \tag{4.124}$$

Unter der Annahme, dass die Raumimpulsantwort $c(t)$ und die Perturbation $p(t)$ statistisch unabhängig und jeweils mittelwertfrei sind, kann die Einhüllende der quadrierten Perturbation $p(t)$ berechnet werden. Ausgehend von (4.124) und (4.120) gilt

$$\begin{aligned} \mathrm{E}\{c^2(t)\} &= \mathrm{E}\{\hat{c}^2(t)\} && (4.125)\\ &= \mathrm{E}\{(\rho c(t) + Ap(t))^2\} && (4.126)\\ &= \mathrm{E}\{\rho^2 c^2(t)\} + \mathrm{E}\{A^2 p^2(t)\}\,. && (4.127) \end{aligned}$$

Durch Einsetzen von (4.124) in (4.127) ergibt sich

$$\mathrm{E}\{A^2 p^2(t)\} = \left(1-\rho^2\right) e^{-2\Delta t}. \tag{4.128}$$

Die essentielle Aussage von (4.128) ist, dass das zeitliche Abklingverhalten der Perturbationen gleich dem Abklingverhalten der Raumimpulsantworten selber ist. Ein entsprechendes Verhalten wurde auch von Zhang et al. experimentell für die bei der blinden Kanalidentifikation gemachten Schätzfehler nachgewiesen [169].

Integration der angenommenen Fehlplatzierung Als weiterer Parameter wird die angenommene Fehlplatzierung in das Modell integriert. Im Folgenden bezeichne t_0 die Zeit, die der Direktschall zum Erreichen der Referenzposition benötigt, die angenommene Entfernung von der Referenzposition wird als D und die Schallgeschwindigkeit mit v bezeichnet. Mit diesen zusätzlichen Parametern kann (4.128) weiter verfeinert werden.

Letztendlich ergibt sich für die Einhüllende der quadrierten Perturbation

$$\mathrm{E}\{p^2(t)\} = \begin{cases} 0, & t < t_0 - \frac{D}{v} \\ 1, & t_0 - \frac{D}{v} \leq t < t_0 + \frac{D}{v} \\ e^{-2\Delta\left(t - t_0 - \frac{D}{v}\right)}, & t \geq t_0 + \frac{D}{v}, \end{cases} \tag{4.129}$$

wobei $\mathrm{E}\{p^2(t)\}$ der Einfachheit halber auf einen Maximalwert von eins normiert wurde.

4.3.2.3 Modellierung der Perturbation im Frequenzbereich

Für die Modellierung der Perturbationen im Frequenzbereich wird das von Radlović et al. erarbeitete Modell verwendet [139, 140]. Die Ergebnisse von Radlović et al. bilden

auch die Grundlage des in Abschnitt 4.3.1 beschriebenen Regularisierungsterms zur Erhöhung der räumlichen Robustheit.

Wie bereits in Abschnitt 3.4.4 beschrieben untersuchten Radlović et al. das Szenario, bei dem ein *perfekter* Entzerrer für die Referenzposition entworfen wurde und dann bei räumlicher Fehlplatzierung angewendet wird.

Die frequenzabhängige Energie der Perturbationen $\mathrm{E}\{|P(\omega)|^2\}$ wird als

$$\mathrm{E}\{|P(\omega)|^2\} = |C(\omega)|^2 \left(2 - 2\frac{\sin(\omega D/v)}{\omega D/v}\right) \tag{4.130}$$

berechnet, wobei $C(\omega)$, $P(\omega)$ sowie $H(\omega)$ die Fourier-Transformierten der Raumimpulsantwort $c(t)$, der additiven Perturbation $p(t)$ sowie des (an der Referenzposition) *perfekten* Entzerrers $h(t)$ sind; D und v bezeichnen die angenommene Distanz zur Referenzposition sowie die Schallgeschwindigkeit.

4.3.2.4 Erzeugung perturbierter Raumimpulsantworten

In den Abschnitten 4.3.2.2 und 4.3.2.3 wurde ein Modell der statistischen Eigenschaften der Perturbationen im Zeit- sowie im Frequenzbereich erarbeitet.

Auf Grundlage von (4.120), (4.129) und (4.130) wird in diesem Abschnitt der Algorithmus zum Generieren der künstlich perturbierten Raumimpulsantworten entwickelt. Diese Methode wurde in [77] publiziert.

Unter der Annahme, dass alle Signale und Impulsantworten unter Einhaltung des Abtasttheorems abgetastet wurden, können die zeitkontinuierlichen Impulsantworten aus (4.120) und (4.129) durch die zeitdiskreten Repräsentationen $c(n)$ und $p(n)$ ersetzt werden; (4.130) wird entsprechend an diskreten Frequenzen ω_k abgetastet.

Nach der folgenden Vorschrift werden auf Grundlage einer gemessenen Raumimpulsantwort R weitere Realisierungen dieser Raumimpulsantwort generiert.

Generierung einer perturbierten Raumimpulsantwort Zur Generierung einer weiteren Realisierung $c^{(r)}(n)$ auf Basis der an der Referenzposition gemessenen Raumimpulsantwort $c(n)$ wird eine Realisierung des additiven Perturbationsterms $p^{(r)}(n)$ generiert. Hierfür werden folgende Schritte ausgeführt:

1. Generierung einer Instanz eines mittelwertfreien, gaußförmigen weißen Rauschens $b^{(r)}(n)$ mit Varianz $\sigma^2_{b(n)} = 1$.

2. Transformation von $b^{(r)}(n)$ in den Frequenzbereich durch die diskrete Fourier-Transformation:

$$b^{(r)}(n) \longleftrightarrow B^{(r)}\left(e^{j\omega_k}\right). \tag{4.131}$$

3. Formung im Frequenzbereich durch Multiplikation mit dem in (4.130) hergeleiteten Frequenzgang:

$$\widehat{B}^{(r)}\left(e^{j\omega_k}\right) = B^{(r)}\left(e^{j\omega_k}\right) \cdot \mathrm{E}\left\{|P(\omega_k)|^2\right\}^{\frac{1}{2}} \tag{4.132}$$

4. Rücktransformation von $\widehat{B}^{(r)}(e^{j\omega_k})$ in den Zeitbereich durch die inverse diskrete Fourier-Transfomation:

$$\hat{b}^{(r)}(n) \longleftrightarrow \widehat{B}^{(r)}\left(e^{j\omega_k}\right). \tag{4.133}$$

5. Multiplikation von $\hat{b}^{(r)}(n)$ mit $\mathrm{E}\{p^2(nT)\}^{\frac{1}{2}}$ aus (4.129), um das gewünschte zeitliche Abklingverhalten zu erzeugen; T bezeichnet hierbei das Abtastintervall ($T = 1/f_s$, wobei f_s die Abtastrate in Hertz ist):

$$\hat{p}^{(r)}(n) = \hat{b}^{(r)}(n) \cdot \mathrm{E}\left\{p^2(nT)\right\}^{\frac{1}{2}}. \tag{4.134}$$

6. Mittelwertbefreiung der Perturbation:

$$p^{(r)}(n) = \hat{p}^{(r)}(n) - \left(\frac{1}{L_c}\sum_{n=0}^{L_c-1}\hat{p}^{(r)}(n)\right). \tag{4.135}$$

Die so erzeugte Realisierung des Störterms $p^{(r)}(n)$ approximiert die in diesem Abschnitt hergeleiteten statistischen Eigenschaften im Zeit- und im Frequenzbereich.

Um letztendlich eine Realisierung der perturbierten Raumimpulsantwort $c^{(r)}(n)$ zu erzeugen, wird die zeitdiskrete Version von (4.120) verwendet:

$$c^{(r)}(n) = \rho c(n) + A p^{(r)}(n). \tag{4.136}$$

Wahl der Faktoren ρ und A Die Faktoren ρ und A in Gleichung (4.136) werden so gewählt, dass die perturbierte Raumimpulsantwort $c^{(r)}(n)$ die gleiche Energie wie die Raumimpulsantwort $c(n)$ an der Referenzposition hat

$$\sum_{n=0}^{L_c-1} c^2(n) = \sum_{n=0}^{L_c-1} {c^{(r)}}^2(n)\,, \tag{4.137}$$

und dass der normalisierte Systemfehler (NSM, [126]) einen vorgeschriebenen Wert erreicht

$$M_{\mathrm{dB}} = -10\log_{10}\left(\frac{\rho^2 \boldsymbol{c}^T\boldsymbol{c}}{A^2 \boldsymbol{p}^{(r)T}\boldsymbol{p}^{(r)}}\right). \tag{4.138}$$

Hierfür wird in einem ersten Schritt ein Faktor $\tilde{A}$ als

$$\tilde{A} = 10^{\frac{10\left(\log_{10}\left(\boldsymbol{c}^T\boldsymbol{c}\right)-\log_{10}\left(\boldsymbol{p}^{(r)T}\boldsymbol{p}^{(r)}\right)\right)-M_{\mathrm{dB}}}{20}} \tag{4.139}$$

bestimmt. In einem zweiten Schritt wird die perturbierte Instanz $\tilde{c}^{(r)}(n)$ der Raumimpulsantwort $c(n)$ als

$$\tilde{c}^{(r)}(n) = c(n) + \tilde{A}p^{(r)}(n) \tag{4.140}$$

generiert; die perturbierte Raumimpulsantwort weist jetzt die gewünschte Fehlerenergie M_{dB} auf. In einem letzten Schritt wird die perturbierte Raumimpulsantwort $\tilde{c}^{(r)}(n)$ aus (4.140) so skaliert, dass sie die gleiche Energie wie die Raumimpulsantwort an der Referenzposition aufweist:

$$\boldsymbol{c}^{(r)} = \left(\frac{\|\boldsymbol{c}\|_2}{\left\|\tilde{\boldsymbol{c}}^{(r)}\right\|_2}\right)\tilde{\boldsymbol{c}}^{(r)}. \tag{4.141}$$

Durch die Hintereinanderausführung von (4.139), (4.140) und (4.141) werden die Faktoren ρ und A aus (4.136) implizit entsprechend der vorgegebenen Fehlerenergie gewählt.

In Abbildung 4.8 ist exemplarisch eine gemessene Raumimpulsantwort (links) und die nach dem in diesem Abschnitt präsentierten Algorithmus generierte Perturbation (rechts) für $D = 2\,\mathrm{cm}$ und $M_{\mathrm{dB}} = -5\,\mathrm{dB}$ dargestellt.

4.4 Wahl der Gewichtsfenster

In diesem Abschnitt wird die Wahl der Gewichtsfenster für die gewünschten und ungewünschten Anteile der globalen Impulsantworten für die gewünschten Signalpfade

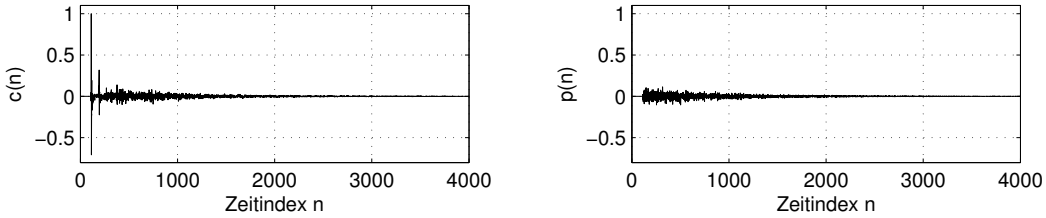

Abbildung 4.8: Eine gemessene Raumimpulsantwort (links) und eine Instanz der nach dem in diesem Abschnitt präsentierten Algorithmus erstellten Perturbationen (rechts) für eine angenommene Fehlplatzierung von $D = 2\,\mathrm{cm}$ und ein gewünschtes NSM von $M_{\mathrm{dB}} = -5\,\mathrm{dB}$.

sowie für die Übersprechkomponenten präsentiert. Die Wahl der Gewichtsfenster kann als Verallgemeinerung der von Mertins et al. in [120] sowie von Mei et al. in [114] vorgeschlagenen Gewichtsfenster verstanden werden.

Die Übersprechkompensation wird in dem entwickelten Formelwerk durch eine entsprechende Wahl der Gewichtsfenster für die gewünschten und die ungewünschten Signalpfade abgebildet [79, 80]. Das Problem der Übersprechkompensation wird so direkt in dem in Abschnitt 4.1.2.1 entwickelten Formelwerk abgebildet. Hierzu wird für jedes der Q Quellensignale $s_q(n)$ festgelegt, an welcher der M Abhörpositionen dieses Quellensignal ein *gewünschtes* Signal ist oder eine *ungewünschte* Übersprechkomponente darstellt [79]; für die alleinige Raumhallkompensation werden alle Signalpfade als *gewünscht* angenommen.

Wahl der Gewichtsfenster für die gewünschten Signalpfade Ist der Pfad von der q-ten Signalquelle zur m-ten Abhörposition ein gewünschter Signalpfad, so werden die von Mertins et al. in [120] vorgeschlagene Gewichtsfenster verwendet [79]. Hierbei wird für den gewünschten Anteil einer globalen Impulsantwort $g_{mq}(n)$ ein Gewichtsfenster $w_{mq}^{(d)}(n)$ derart gewählt, dass die ersten 4 Millisekunden beginnend mit der Direktschallkomponente selektiert werden. Die Gewichtsfunktion für den gewünschten Teil eines Signalpfades ist durch

$$w_{mq}^{(d)}(n) = \begin{cases} 1, & t_{0m} f_s \leq n < (t_{0m} + T_d)\, f_s, \\ 0, & \text{sonst}, \end{cases} \tag{4.142}$$

gegeben, wobei f_s die Abtastrate und $T_d = 4\,\mathrm{ms}$; t_{0m} ist die durchschnittliche Laufzeit der Direktschallkomponente aller N Lautsprecher zur m-ten Mikrofonposition.

Für den ungewünschten Anteil der globalen Impulsantwort wird, wie in [120] vorgeschlagen, das Reziproke der durchschnittlichen temporalen Maskierungsschwelle des menschlichen auditorischen Systems nach [35] verwendet. Das Gewichtsfenster wird als

$$w_{mq}^{(u)} = l_m(n) \tag{4.143}$$

mit

$$l_m(n) = \begin{cases} 10^{\frac{3}{\log(N_{0m}/(N_{1m}+N_2))}\log\left(\frac{n}{N_{1m}+N_2}\right)+0.5}, & (N_{1m}+N_2) \leq n < L_g, \\ 0, & \text{sonst}, \end{cases} \tag{4.144}$$

und $N_{1m} = t_{0m} f_s$, $N_2 = T_d f_s$ sowie $N_{0m} = (0.2\,\mathrm{s} + t_{0m})\, f_s$ definiert [79].

4.4.1 Übersprechkompensation

Eine Lösung für die Übersprechkompensation mit zwei Quellensignalen, zwei Lautsprechern sowie zwei Abhörpositionen (siehe Abbildung 3.23) bei gleichzeitiger Raumhallkompensation durch Formung der globalen Impulsantworten durch die Optimierung einer auf der p-Norm basierenden Zielfunktion wurde erstmalig im Rahmen dieser Arbeit entwickelt und in [80] veröffentlicht. In [80] wurden auf Grundlage mehrerer gemessener Raumimpulsantworten (nach dem Prinzip von Mei et al. [114]) robuste Filter für die gemeinsame Raumhallkompensation und Übersprechkompensation entworfen.

Wahl der Gewichtsfenster für die Übersprechkomponenten Das Gewichtsfenster für den ungewünschten Anteil der globalen Impulsantwort eines Übersprech-Pfades ist als

$$w_{mq}^{(u)}(n) = \max\Big\{w_0, \max_m\{l_m(n)\}\Big\} \tag{4.145}$$

definiert, wobei $l_m(n)$ in (4.144) für die m-te Abhörposition gegeben ist [79]. Der Parameter w_0 in (4.145) bildet die gewünschte Unterdrückung der Übersprechkomponente im Vergleich zum gewünschten Signal ab (vergleiche (3.53)).

Die Wahl dieses Gewichtsfensters für die Übersprech-Pfade garantiert, dass der Nachhall der Übersprechkomponente nicht den Nachhall der gewünschten Signalkomponente übersteigt [79, 80].

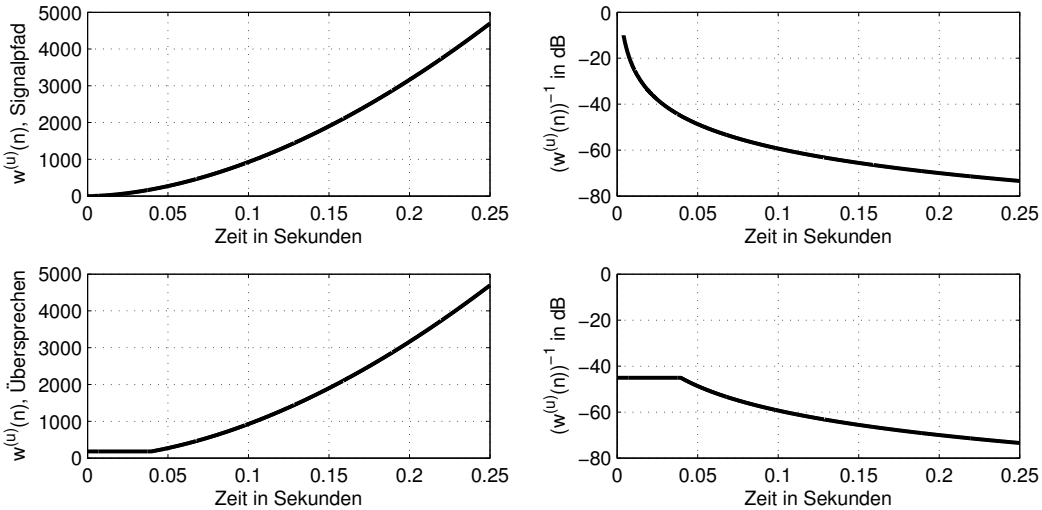

Abbildung 4.9: Beispiele für die Gewichtsfenster. In der oberen Zeile ist das Gewichtsfenster für einen Signalpfad auf einer linearen Skala (links) sowie das Reziproke des Gewichtsfensters, welches die temporale Maskierungsschwelle approximiert, auf einer logarithmischen Skala (rechts) dargestellt. In der unteren Zeile ist das Gewichtsfenster für den ungewünschten Teil eines Übersprech-Pfades mit $w_0 = 180$ auf einer linearen Skala (links) und das entsprechende Reziproke auf einer logarithmischen Skala (rechts) dargestellt.

Da es für die Übersprechkomponente keinen *gewünschten* Anteil gibt wird das Gewichtsfenster für den gewünschten Teil der globalen Impulsantworten der Übersprechpfade als

$$w_{mq}^{(d)}(n) = 0 \tag{4.146}$$

definiert.

In Abbildung 4.9 sind die Gewichtsfenster für die ungewünschten Anteile der globalen Impulsantworten für die gemeinsame Enthallung und Übersprechkompensation exemplarisch dargestellt. Das Reziproke der Gewichtsfenster entspricht hierbei jeweils der durchschnittlichen temporalen Maskierungsschwelle des menschlichen auditorischen Systems nach [35].

4.5 Zusammenfassung

In diesem Abschnitt wurden die im Rahmen dieser Arbeit entwickelten Methoden und Algorithmen präsentiert. Es wurde ein verallgemeinertes Formelwerk für beliebige Übertragungsszenarien zur Raumhallkompensation in halligen Umgebungen entwickelt.

Des Weiteren wurden zwei neue Methoden zum Entwurf von räumlich robusten Entzerrern erarbeitet. Zum einen wurde die Veränderung der Raumimpulsantwort bei einer Fehlplatzierung statistisch beschrieben und als Regularisierungsterm in das Optimierungsproblem zum Filterentwurf integriert. Als zweiter Ansatz wurde ein Verfahren zur Generierung von künstlich perturbierten Raumimpulsantworten entwickelt. Die künstlich perturbierten Raumimpulsantworten werden als zusätzliche Realisierungen in den Filterentwurf einbezogen und ersetzen somit zusätzliche Messungen.

Darüber hinaus wurden zwei neue Methoden zur Formung der globalen Impulsantworten im Zeitbereich und zur gleichzeitigen Entzerrung des Frequenzgangs entwickelt. Beide Methoden basieren auf der Integration eines zusätzlichen Regularisierungsterms in das Optimierungsproblem zum Filterentwurf. Die erste Methode basiert analog zur Formung im Zeitbereich auf einem Kriterium auf Basis der p-Norm. Die zweite entwickelte Methode verwendet eine generalisierte Form des Maßes zur Bewertung des spektralen Flachheit.

Abschließend wurde gezeigt, wie das Problem der Übersprechkompensation durch die Wahl geeigneter Gewichtsfenster innerhalb des erarbeiteten Formelwerks gelöst werden kann.

Die im Rahmen dieser Arbeit entwickelten Methoden und Verfahren wurden in Teilen in [72–81] veröffentlicht.

5 Daten

In diesem Kapitel wird der für die Experimente verwendete Datensatz beschrieben. In Abschnitt 5.1 werden die zur Messung der Raumimpulsantworten verwendeten Hardwarekomponenten vorgestellt. Der Messaufbau sowie die Durchführung der Messungen werden in den Abschnitten 5.2 und 5.3 beschrieben.

5.1 Verwendete Komponenten

Zur Untersuchung der räumlichen Robustheit der entworfenen Filter sowie zur Untersuchung der Übersprechkompensation wurde ein Satz an Raumimpulsantworten im Seminarraum des Instituts für Signalverarbeitung der Universität zu Lübeck aufgenommen.

Für die Reproduktion der Messsignale im Seminarraum wurden Klein&Hummel M52 Lautsprecher verwendet. Die Aufnahme erfolgte mit einem Kunstkopf. Als Kunstkopf wurde ein Cortex MK-2 verwendet, der mit MK250 Mikrofonkapseln ausgestattet wurde. Die Lautsprecher sowie die Mikrofone wurden mit einem AD/DA Wandler vom Typ RME Fireface800 angesteuert; dieses erlaubt die synchronisierte Wiedergabe und Aufnahme von bis zu acht Kanälen. Abbildung 5.1 zeigt Fotos des Lautsprechers, des Kunstkopfes sowie des AD/DA Wandlers.

Der Seminarraum verfügt über keinerlei akustische Dämpfung und hat eine Größe von $[6.85 \times 5.3 \times 3]\ \mathrm{m}^3$. Abbildung 5.2 zeigt ein Foto des Seminarraums.

Die Nachhallzeit des Seminarraumes wurde mit Hilfe der Rückintegrations-Methode nach [146] auf $T_{60} = 380$ Millisekunden geschätzt.

5.2 Messaufbau

Die Lautsprecher und der Kunstkopf wurden wie in Abbildung 5.3 dargestellt in dem Seminarraum aufgebaut. Die vier Lautsprecher waren $1.6\,\mathrm{m}$ von der Wand entfernt und in einer Höhe von $1.5\,\mathrm{m}$ über dem Fussboden positioniert. Die Ohren des Kunstkopfs waren

Abbildung 5.1: Die zur Messung im Seminarraum verwendeten Komponenten: Lautsprecher Klein&Hummel M52 (links), Kunstkopf Cortex MK-2 (rechts oben) sowie das RME Fireface800 (rechts unten).

$1.6\,\mathrm{m}$ über dem Boden positioniert. Der Kunstkopf wurde auf einer Positioniereinheit mit einer Positionsgenauigkeit im Submillimeterbereich montiert. Die einzelnen Positionen innerhalb des Abhörvolumens wurden mit Hilfe der Positioniereinheit sukzessive angefahren.

Nachdem ein Abhörvolumen gemessen wurde, wurde der Kunstkopf $0.8\,\mathrm{m}$ neben der ersten Referenzposition aufgestellt, dies ist durch die beiden Kunstköpfe in Abbildung 5.3 an den Positionen P_1 und P_2 angedeutet. An dieser Position wurde ein zweites Abhörvolumen abgetastet, um die Übersprechkompensation mit zwei Zuhörern, also mit mehr als zwei Signalsenken, zu untersuchen.

5.3 Durchführung

Für die Messung der Raumimpulsantworten wurden *exponential sine sweeps* als Anregungssignal benutzt [32]. Diese Methode wurde in Abschnitt 2.1.3.2 erklärt.

Am Institut für Signalverarbeitung der Universität zu Lübeck wurde eine MATLAB Toolbox entwickelt, die die Messung von Raumimpulsantworten mit unterschiedlichen

Abbildung 5.2: Der Seminarraum in dem die Messungen durchgeführt wurden.

Anregungssignalen und frei wählbaren Parametern kapselt. Ein Bildschirmfoto der Benutzeroberfläche der Toolbox ist in Abbildung 5.4 dargestellt.

Für die Messung wurde eine Abtastrate von $f_s = 48\,\mathrm{kHz}$ verwendet. Als Parameter für den *exponential sine sweep* wurden $f_L = 20\,\mathrm{Hz}$ als untere und $f_U = 22\,\mathrm{kHz}$ als obere Grenzfrequenz verwendet. Die Länge des Messsignals betrug $T = 4\,\mathrm{s}$. Die Messung und Rekonstruktion der Raumimpulsantworten wurde wie in [32] beschrieben durchgeführt. Nach der Messung wurden die Raumimpulsantworten auf $f_s = 16\,\mathrm{kHz}$ heruntergetastet und auf eine Länge von $L_c = 4000$ Zeitkoeffizienten beschränkt.

Die durch das Außenohr verursachten spektralen Verzerrungen wurden durch ein lineares Prädiktorfehlerfilter der Länge $L_p = 11$ kompensiert.

Eine der so gemessenen Raumimpulsantworten ist in Abbildung 5.5 auf einer logarithmischen Skala dargestellt; des Weiteren ist der Betragsfrequenzgang dargestellt. Diese Raumimpulsantwort wird exemplarisch für die Experimente zur einkanaligen Entzerrung verwendet.

Um das Abhörvolumen abzutasten, wurde der Kunstkopf in einem $[2\times2\times2]\ \mathrm{cm}^3$ großen Volumen bewegt. Das räumliche Abtastintervall betrug $\Delta = 1\,\mathrm{cm}$ in jede Raumdimension und erfüllt somit das Abtasttheorem für Raumimpulsantworten (Abschnitt 3.4.5

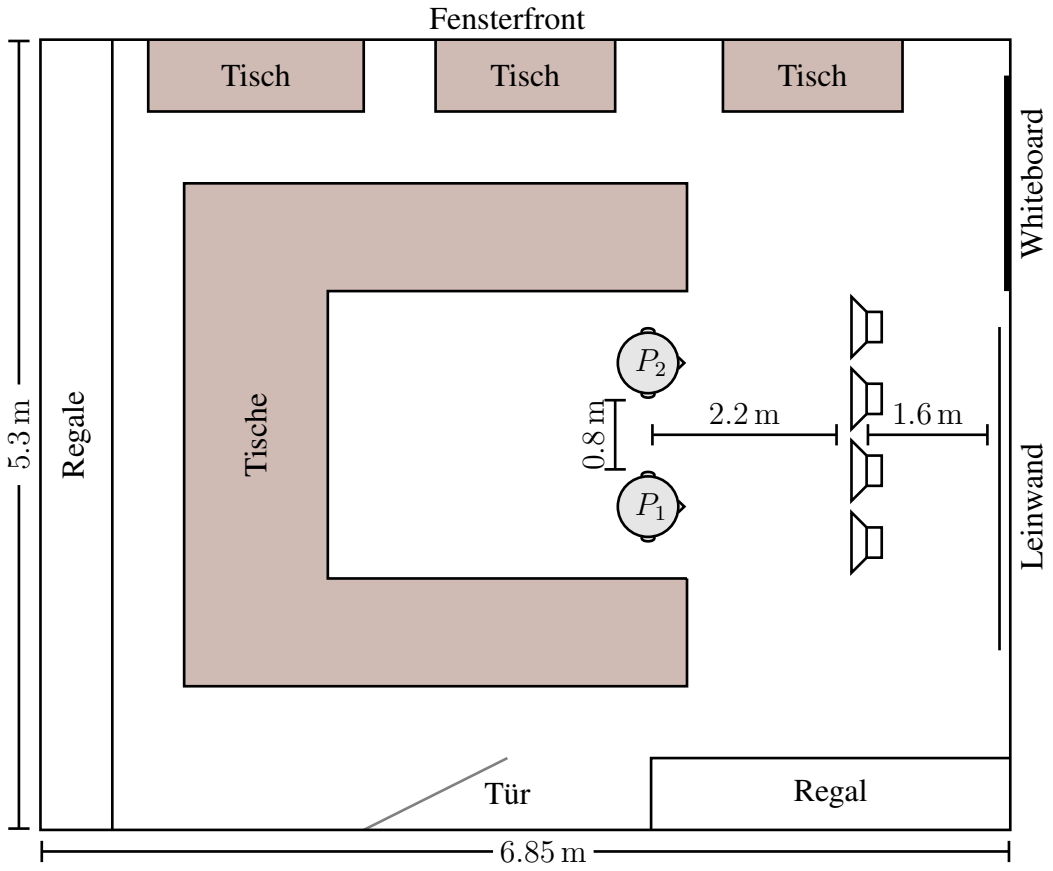

Abbildung 5.3: Schematische Darstellung des Messaufbaus im Seminarraum. Die Position der Kunstköpfe wird mit P_1 beziehungsweise P_2 bezeichnet.

sowie [1, 114]); es wurden also insgesamt 27 Raumimpulsantworten pro Abhörvolumen gemessen.

Des Weiteren wurden 40 zusätzliche Raumimpulsantworten innerhalb des Abhörvolumens gemessen, die jedoch disjunkt mit den Referenzpositionen waren. Diese zusätzlichen Messungen werden zur Evaluierung der Entzerrung verwendet. In Abbildung 5.6 ist grafisch die Verteilung der Messpositionen innerhalb des Abhörvolumens dargestellt.

Anzahl der Messungen Insgesamt wurden 108 Raumimpulsantworten an den Referenzpositionen (zwei Ohren, zwei Positionen, jeweils 27 Messpunkte) sowie 160 Raumimpulsantworten an den Testpositionen (zwei Ohren, zwei Positionen, jeweils 40 Messpunkte) gemessen.

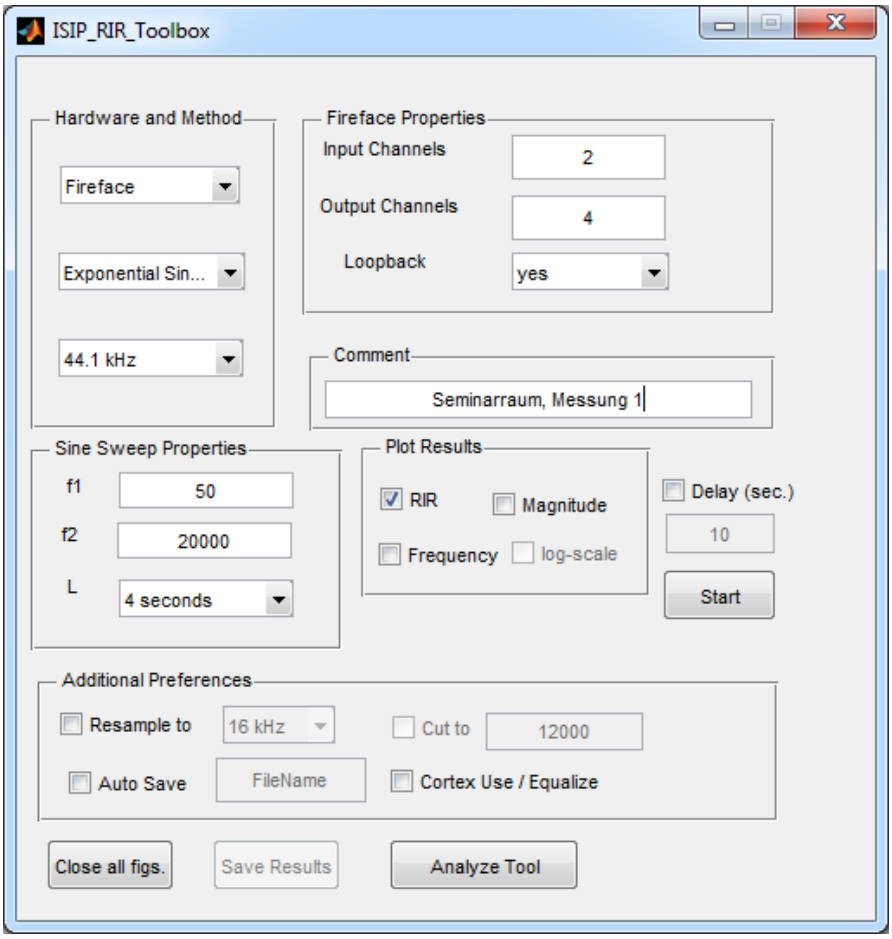

Abbildung 5.4: Benutzeroberfläche der ISIP RIR-Toolbox. Mit Hilfe dieser Toolbox wurden die Messungen der Raumimpulsantworten durchgeführt.

5.4 Weitere Raumimpulsantworten

Für die im Rahmen dieser Arbeit durchgeführten Experimente wurden zwei weitere gemessene Raumimpulsantworten verwendet.

Messung im Technologie- und Gründerzentrum Oldenburg Eine weitere Raumimpulsantwort wurde im Technologie- und Gründerzentrum in Oldenburg (TG Oldenburg) in einem Toilettenraum im Fraunhofer Trakt gemessen. Als Anregungssignal wurde ein *exponential sine sweep* [32] verwendet. Die Raumimpulsantwort wurde freundlicherweise von Dr.-Ing. Stefan Goetze zur Verfügung gestellt; sie wurde mit einer Abtastrate von $f_s = 16\,\text{kHz}$ gemessen und auf $L_c = 16000$ Zeitkoeffizienten begrenzt.

Die Raumimpulsantwort sowie der Betragsfrequenzgang sind in Abbildung 5.7 jeweils auf einer logarithmischer Skala dargestellt.

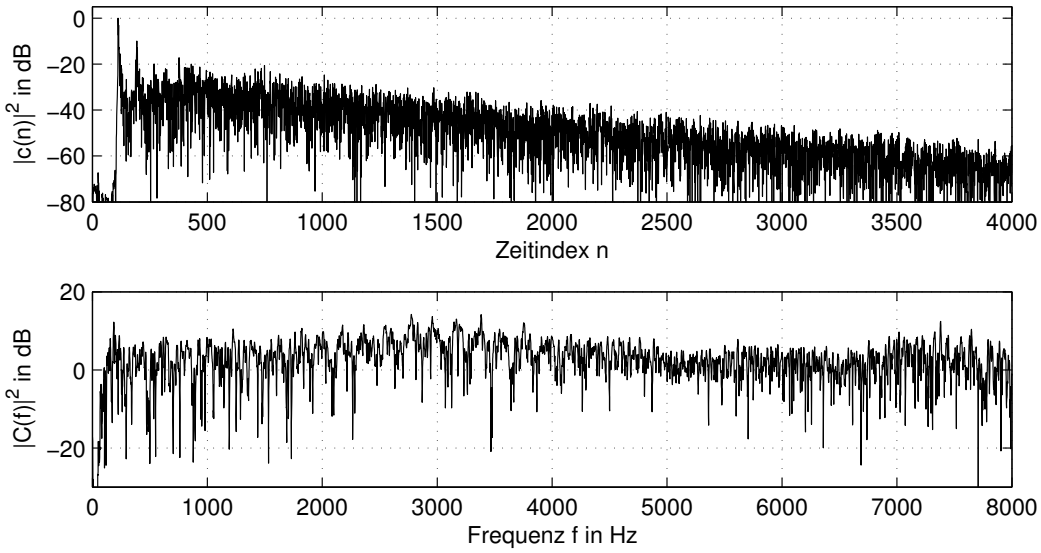

Abbildung 5.5: Eine Raumimpulsantwort aus dem mit dem Kunstkopf gemessenen Datensatz auf linearer (oben) und logarithmischer (mitte) Skala sowie der Betragsfrequenzgang (unten). Diese Raumimpulsantwort wird exemplarisch für die Untersuchungen der einkanaligen Methoden zur Entzerrung verwendet.

Messung im Wohn- und Esszimmer Eine weitere Raumimpulsantwort wurde in einem Wohn- und Esszimmer mit 44 Quadratmetern Grundfläche gemessen. Zur Wiedergabe des Messsignals (ein *exponential sine sweep* [32]) wurde eine hochwertige HiFi Anlage verwendet. Zur Messung wurde ein Beyerdynamics MM1 Mikrofon verwendet. Die Abtastrate betrug $f_s = 44.1\,\mathrm{kHz}$ und die Raumimpulsantwort wurde auf $L_c = 8000$ Zeitkoeffizienten begrenzt. Die Messung wurde von Prof. Dr.-Ing. Alfred Mertins durchgeführt und zur Verfügung gestellt. Die gemessene Raumimpulsantwort sowie der Betragsfrequenzgang sind in Abbildung 5.8 dargestellt.

Im Betragsfrequenzgang ist deutlich der Effekt des Tiefpassfilters zur Rekonstruktion des Digital/Analog-Wandlers zu erkennen.

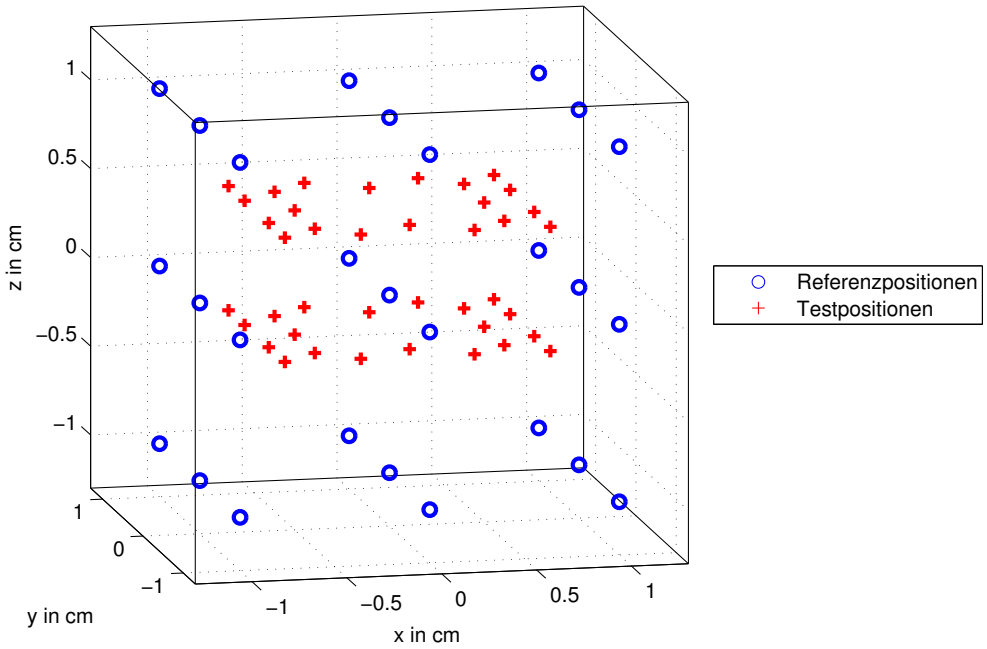

Abbildung 5.6: Grafische Darstellung der Verteilung der Messpunkte innerhalb des $[2\times2\times2]$ cm^3 großen Abhörvolumens.

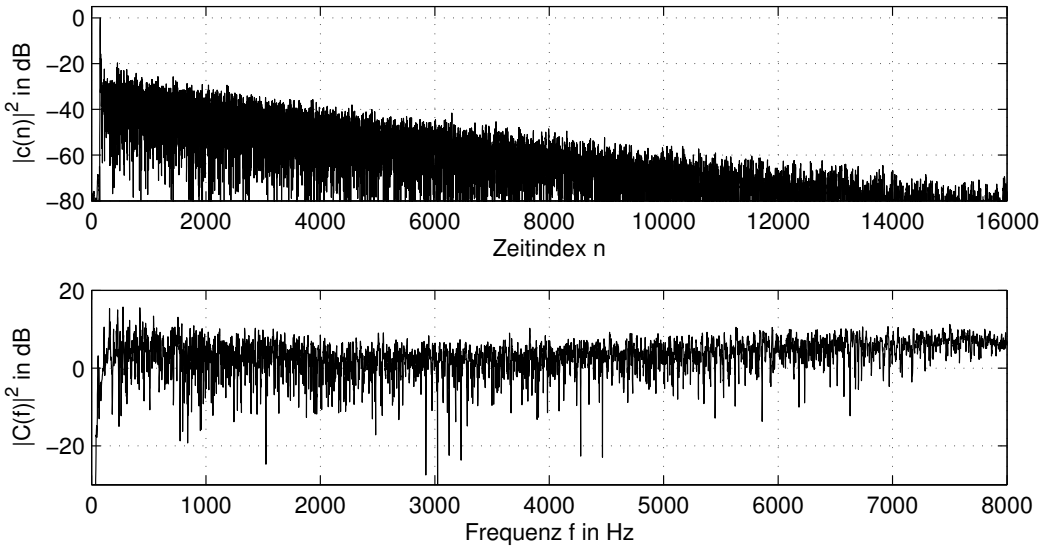

Abbildung 5.7: In einem Toilettenraum des Technologie- und Gründerzentrums in Oldenburg gemessene Raumimpulsantwort auf einer logarithmischen Skala (oben) sowie der quadrierte Betragsfrequenzgang (unten).

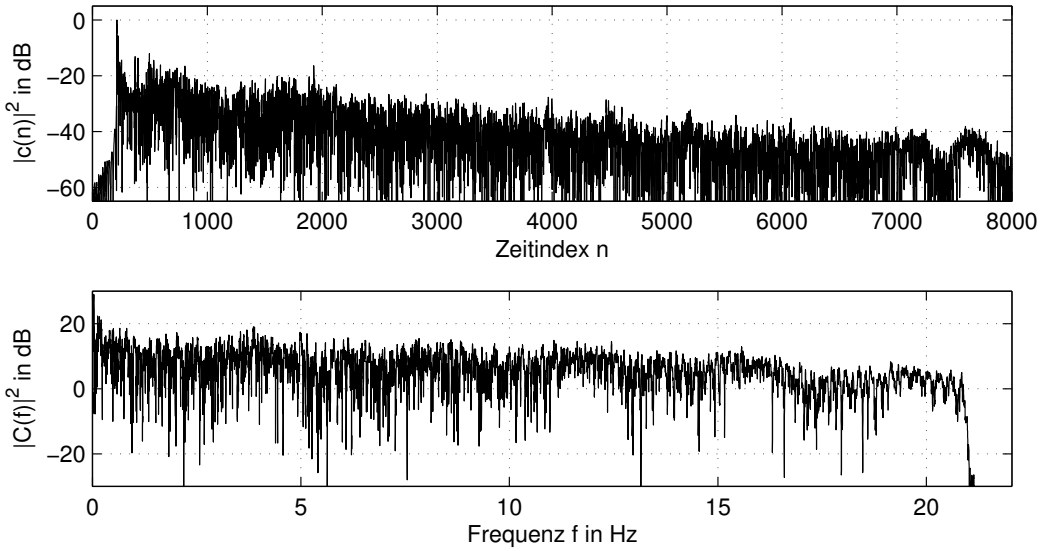

Abbildung 5.8: In einem Wohn- und Esszimmer mit einer hochwertigen HiFi Anlage gemessene Raumimpulsantwort auf einer logarithmischen Skala (oben). Im Betragsfrequenzgang (unten) ist deutlich der Effekt des Rekonstruktionstiefpassfilters des D/A Wandlers zu erkennen.

6 Experimente und Ergebnisse

In diesem Kapitel werden die mit den neu entworfenen Methoden erzielten Ergebnisse präsentiert. Des Weiteren werden zur quantitativen Evaluierung der Ergebnisse zwei neue Maße eingeführt, die den wahrnehmbaren Nachhall sowie die Übersprechkompensation quantisieren. Die Maße werden in Abschnitt 6.1 eingeführt.

Zur besseren Übersicht werden die erzielten Ergebnisse am Ende jedes Abschnitts kurz bewertet.

6.1 Neue Maße zur objektiven Bewertung der Ergebnisse

Die üblicherweise zur Quantifizierung von akustischen Kanälen verwendeten Maße (siehe Abschnitt 2.3, sowie [44, 46, 98]) sind nicht geeignet, um die akustischen Kanäle nach Formung durch ein auf der p-Norm basierendes Kriterium zu bewerten. So erfassen die üblicherweise in der Literatur verwendeten Maße in den meisten Fällen die Verteilung der Energie über die Zeitkoeffizienten, also ein quadratisches Kriterium [44, 98].

6.1.1 Quantifizierung des wahrnehmbaren Nachhalls

Im Rahmen dieser Arbeit wurde das *normalized perceivable reverberation quantization measure* (kurz nPRQ) Maß erarbeitet und in [79] publiziert, welches die normalisierte Variante des in [81] eingeführten *perceivable reverberation quantization measures* (kurz pRQ) ist; das pRQ Maß basiert wiederum auf den Arbeiten von Mei et al. [114]. In [42] wurde durch Hörtests gezeigt, dass das pRQ Maß positiv mit der Intensität des wahrnehmbaren Nachhall korreliert ist.

Das nPRQ Maß basiert auf Plausibilitätsüberlegungen und setzt das raumakustische Maß der Nachhallzeit T_{60} [98] mit dem psychoakustischen Konzept der temporalen Maskierung anhand der durchschnittlichen temporalen Maskierungsschwelle nach Fielder [35] in Verbindung. Das nPRQ Maß ist der Wert des durchschnittlichen Überschreitens

der quadrierten Zeitkoeffizienten über die temporale Maskierungsschwelle und über $-60\,\mathrm{dB}$.

Die Formel zur Berechnung des nPRQ Maßes lautet

$$\mathrm{nPRQ} = \begin{cases} \frac{1}{\|\boldsymbol{g}_E\|_0} \cdot \sum\limits_{n=N_0}^{L_g-1} g_E(n)\,, & \text{für} \quad \|\boldsymbol{g}_E\|_0 > 0 \\ 0, & \text{sonst,} \end{cases} \tag{6.1}$$

wobei $\|\cdot\|_0$ die ℓ_0-Pseudonorm ist und $\boldsymbol{g}_E = [g_E(0)\,, \ldots, g_E(L_g-1)]^T$ durch

$$g_E(n) = \begin{cases} 20\log_{10}(|g(n)| \cdot w_u(n))\,, & \text{für} \quad |g(n)| > \max\left\{\frac{1}{w_u(n)}, -60\,\mathrm{dB}\right\}, \\ 0, & \text{sonst,} \end{cases} \tag{6.2}$$

gegeben ist [79].

In Abbildung 6.1 ist die Bedeutung der Gleichungen (6.1) und (6.2) grafisch dargestellt. Der Wert des nPRQ Maßes entspricht der durchschnittlichen Energie des wahrnehmbaren Nachhalls, also des roten Bereichs in Abbildung 6.1.

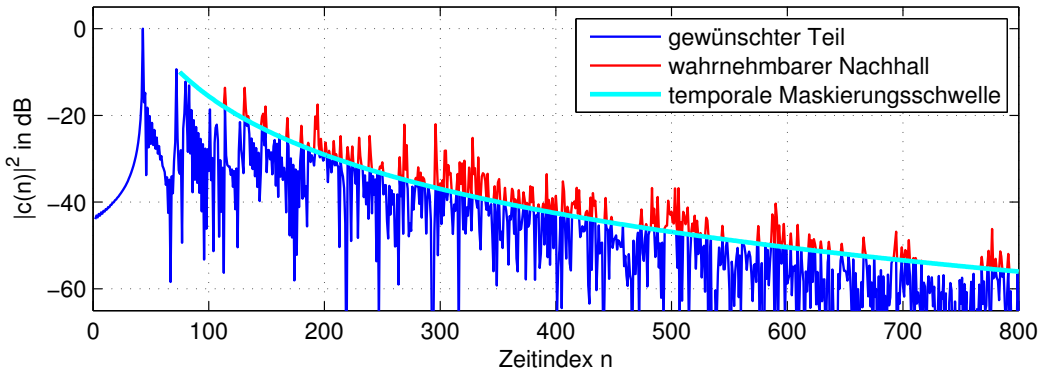

Abbildung 6.1: Das nPRQ Maß ist das durchschnittliche Überschreiten der temporalen Maskierungsschwelle auf einer logarithmischen Skala und über $-60\,\mathrm{dB}$. Das nPRQ Maß kann somit als Mittelung über den *roten* Bereich verstanden werden. Für diese Raumimpulsantwort gilt $\mathrm{nPRQ} = 3.72\,\mathrm{dB}$.

Im Falle einer vollständigen Enthallung durch Formung der Raumimpulsantwort liegt entweder kein Zeitkoeffizient über der temporalen Maskierungsschwelle oder aber die Energie der Zeitkoeffizienten liegt unter $-60\,\mathrm{dB}$; in beiden Fällen gilt $\mathrm{nPRQ} = 0$. Abbildung 6.2 zeigt die Raumimpulsantwort aus Abbildung 6.1 nach der Formung. In

diesem Beispiel konnte das nPRQ Maß durch die Formung nach [120] von $3.72\,\mathrm{dB}$ auf null reduziert werden.

Erste informelle Hörtests bestätigen die Eignung des nPRQ Maßes zur Quantisierung des wahrnehmbaren Nachhalls.

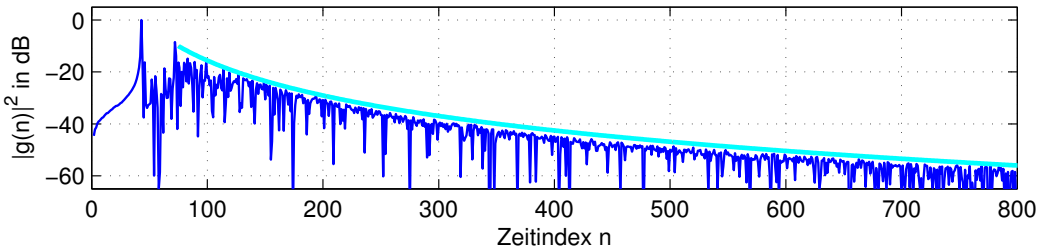

Abbildung 6.2: Vollständig geformte Raumimpulsantwort. Durch die Formung ist kein wahrnehmbarer Nachhall mehr vorhanden und das nPRQ Maß ist null.

6.1.2 Bewertung der Übersprechkompensation

Um die Kompensation des Übersprechens zu quantifizieren wird im Rahmen dieser Arbeit der Wert des betragsmäßig größten Zeitkoeffizienten des gewünschten Signalpfades $g_s(n)$ mit dem Wert des betragsmäßig größten Zeitkoeffizienten des Übersprechpfades $g_c(n)$ auf einer logarithmischen Skala in Relation gesetzt. Dieses Maß wurde in [80] und [79] publiziert und in [79] als *direct signal to crosstalk ratio* (kurz DSCR) bezeichnet.

In Abbildung 6.3 ist das DSCR Maß an einem Beispiel erläutert; in diesem Fall beträgt das DSCR Maß $38\,\mathrm{dB}$.

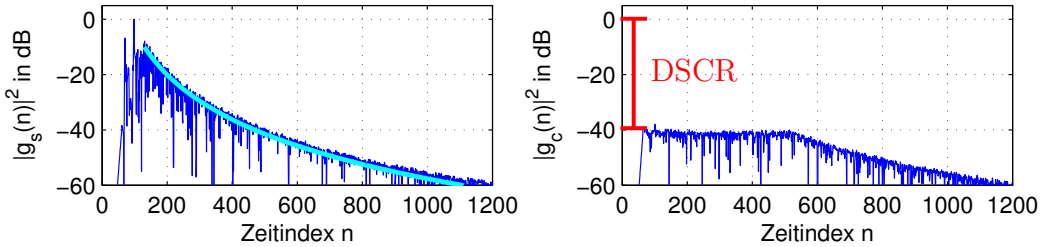

Abbildung 6.3: Das DSCR Maß quantifiziert den Grad der Übersprechkompensation. Hierzu wird der betragsmäßig größte Zeitkoeffizient des gewünschten Signalpfades (links) mit dem betragsmäßig größten Zeitkoeffizienten des Übersprechpfades (rechts) in Relation gesetzt. In diesem Beispiel gilt $\mathrm{DSCR} = 38\,\mathrm{dB}$. Die hellblaue Linie markiert die durchschnittliche temporale Maskierungsschwelle nach Fielder [35].

6.2 Ergebnisse

In diesem Abschnitt werden die mit den neu erarbeiteten Algorithmen erzielten Ergebnisse präsentiert. Als Datengrundlage dienen hierbei die in Kapitel 5 beschriebenen Messungen.

In Abschnitt 6.2.1 werden die Ergebnisse der Impulsformung im Zeitbereich bei gleichzeitiger Optimierung der Frequenzbereichsdarstellung präsentiert. Hierzu wird exemplarisch eine der mit dem Kunstkopf gemessenen Raumimpulsantworten sowie die beiden in Abschnitt 5.4 vorgestellten Raumimpulsantworten verwendet. Des Weiteren wird demonstriert, wie mit dem auf dem generalisierten Maß zur Bewertung der spektralen Flachheit (gSFM) basierenden Verfahren (siehe Abschnitt 4.2.2.1 sowie 4.2.2.2) ein Gesamtsystem mit vorgegebenem Betragsfrequenzgang erzielt werden kann.

In Abschnitt 6.2.2 werden die mit den in Abschnitt 4.3 erarbeiteten Methoden zur räumlich robusten Raumhallkompensation erzielten Ergebnisse vorgestellt. Hierfür wird der mit dem Kunstkopf aufgenommene Datensatz verwendet.

Die Ergebnisse für die Übersprechkompensation in dem erarbeiteten Formelwerk werden in Abschnitt 6.2.3 präsentiert. Die Filter zur Übersprechkompensation wurden für den mit dem Kunstkopf aufgenommenen Datensatz entworfen.

Die Optimierung der jeweiligen Zielfunktion wurde mit einem iterativen Gradientenabstiegsverfahren [14] durchgeführt. Hierbei wurden jeweils $\mathrm{NUM_ITER} = 16000$ Iterationen des Gradientenabstiegsverfahrens durchgeführt.

6.2.1 Optimierung der Frequenzdarstellung

Für die Untersuchung der nicht robusten einkanaligen Methoden zur Formung und Entzerrung wurden die Filter mit der gleichen Länge wie die Raumimpulsantworten entworfen, es gilt also jeweils $L_h = L_c$. Für alle Experimente gilt $p_d = 20$, $p_u = 10$. Als Gewichtsfenster für die Zeitbereichsdarstellung wurden die in [120] empfohlenen Gewichtsfenster verwendet.

6.2.1.1 Regularisierung auf Basis der p-Norm

In diesem Abschnitt werden die Ergebnisse bei gleichzeitiger Optimierung der p-Norm der Zeitbereichs- sowie der Frequenzbereichsdarstellung präsentiert. Hierzu wird das in Gleichung (4.34) gegebene Optimierungsproblem gelöst. Die Ergebnisse sind jeweils in tabellarischer Form dargestellt. In den Tabellen bezeichnet α den Gewichtsfaktor für den Regularisierungsterm, p_f bezeichnet die p-Norm des Regularisierungsterms (4.35)

und EQ beziehungsweise GIR notiert, ob der Betragsfrequenzgang des Entzerrers oder der globalen Impulsantwort zur Berechnung der p-Norm verwendet wurde (4.36).

Mit dem Kunstkopf gemessene RIR Für die mit dem Kunstkopf gemessene Raumimpulsantwort (Abbildung 5.5) gilt $\mathrm{nPRQ} = 11.6\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.58$. Für die Formung ohne Regularisierung (also $\alpha = 0$) gilt $\mathrm{nPRQ} = 2.1\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.24$; die geformte globale Impulsantwort ist in Abbildung 6.4 dargestellt.

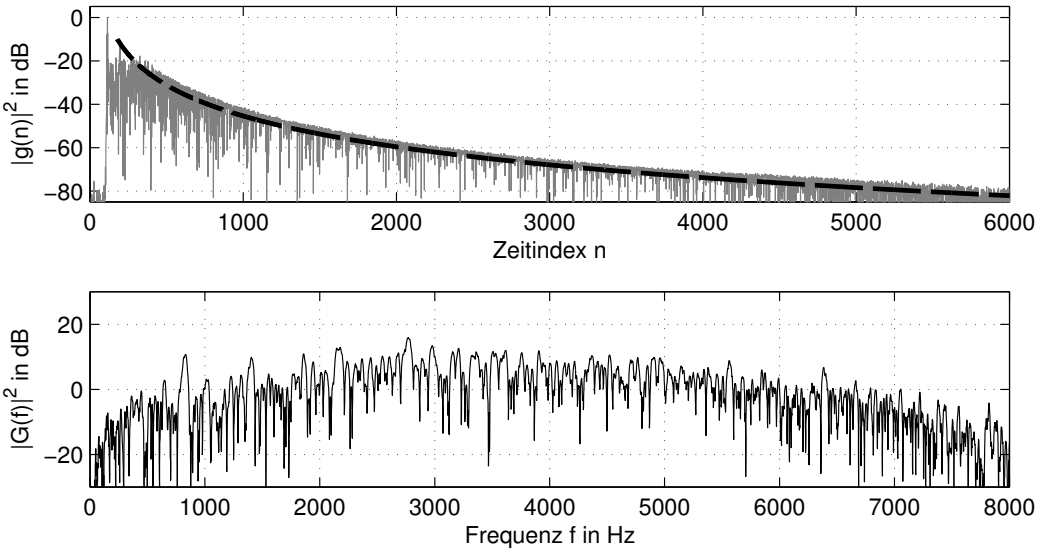

Abbildung 6.4: Ergebnis der Formung der mit dem Kunstkopf gemessenen Raumimpulsantwort ohne Regularisierung. Der Nachhall der geformten Impulsantwort beträgt $\mathrm{nPRQ} = 2.1\,\mathrm{dB}$, die spektrale Flachheit beträgt $\mathrm{SFM} = 0.24$.

Zum Vergleich mit den neu entwickelten Methoden ist in Abbildung 6.5 das Ergebnis bei der Nachfilterung nach Kallinger et al. [85] dargestellt. Das lineare Prädiktorfehlerfilter der Länge $L_p = 20$ wurde für das Gesamtsystem der Formung ohne Regularisierung (Abbildung 6.4) entworfen. Nach der Nachfilterung gilt $\mathrm{nPRQ} = 8.8\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.53$.

In Tabelle 6.1 ist der wahrnehmbare Nachhall, quantifiziert durch das nPRQ Maß aus [79], und in Tabelle 6.2 die spektrale Flachheit für unterschiedliche Parameter der neuen Methode zum Filterentwurf angegeben. Neben der p-Norm der Zeitbereichsdarstellung wurde zusätzlich die p-Norm des Betragsfrequenzgangs optimiert.

In Abbildung 6.6 ist die globale Impulsantwort bei Optimierung der p-Norm des Betragsfrequenzgangs der globalen Impulsantwort ($\alpha = 5$, $p_f = 8$) dargestellt. Der

Setup \ α	Kunstkopf, nPRQ in dB, p-Norm Regularisierung									
	0.1	0.25	0.5	0.75	1	2.5	5	10	20	40
$p_f = 8$, EQ	5.8	6.5	7.3	7.8	8.3	9.6	10.5	11.5	11.5	11.6
$p_f = 8$, GIR	2.1	2.1	2.2	2.2	2.2	2.5	5.9	6.5	8.3	8.0
$p_f = \infty$, EQ	7.0	8.1	9.3	11.9	11.7	11.6	11.6	11.6	11.6	11.6
$p_f = \infty$, GIR	2.3	5.7	6.2	6.4	6.6	7.5	8.1	8.7	9.6	11.7

Tabelle 6.1: Wahrnehmbarer Nachhall, quantifiziert in nPRQ [79]. Als Raumimpulsantwort wurde eine mit dem Kunstkopf gemessene Impulsantwort verwendet. Für die nicht geformte Impulsantwort gilt $\mathrm{nPRQ} = 11.6\,\mathrm{dB}$, für die Formung ohne Regularisierung ergibt sich $\mathrm{nPRQ} = 2.1\,\mathrm{dB}$.

Setup \ α	Kunstkopf, SFM, p-Norm Regularisierung									
	0.1	0.25	0.5	0.75	1	2.5	5	10	20	40
$p_f = 8$, EQ	0.50	0.56	0.59	0.60	0.60	0.60	0.59	0.59	0.59	0.59
$p_f = 8$, GIR	0.25	0.26	0.26	0.27	0.28	0.31	0.62	0.69	0.75	0.79
$p_f = \infty$, EQ	0.59	0.61	0.61	0.61	0.59	0.59	0.58	0.58	0.58	0.58
$p_f = \infty$, GIR	0.25	0.52	0.57	0.59	0.60	0.63	0.64	0.65	0.64	0.59

Tabelle 6.2: Spektrale Flachheit (SFM) nach der Formung. Als Raumimpulsantwort wurde eine mit dem Kunstkopf gemessene Impulsantwort verwendet. Für die nicht geformte Impulsantwort gilt $\mathrm{SFM} = 0.58$, für die Formung ohne Regularisierung ergibt sich $\mathrm{SFM} = 0.24$.

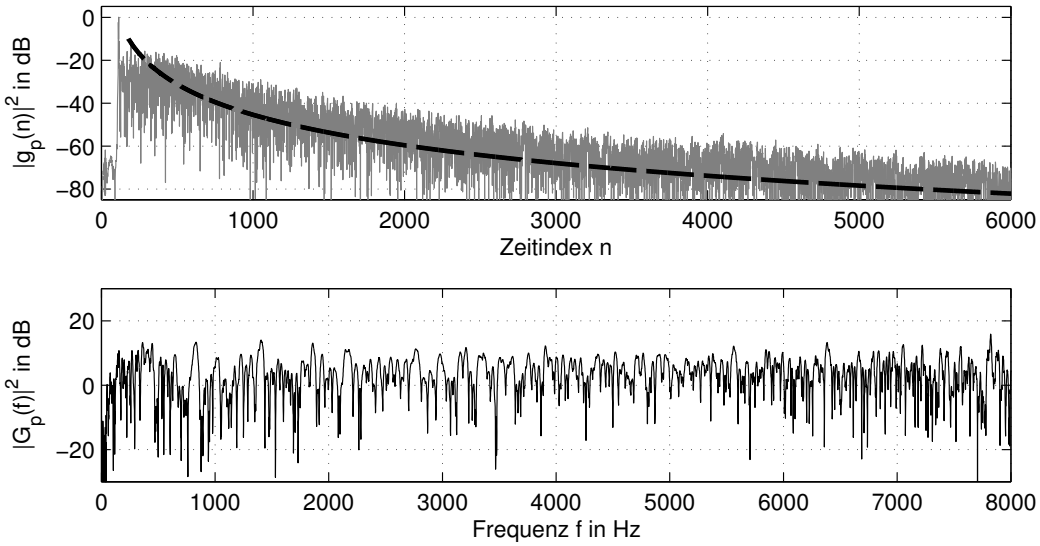

Abbildung 6.5: Ergebnis der Formung der mit dem Kunstkopf gemessenen Raumimpulsantwort ohne Regularisierung. Der Betragsfrequenzgang der geformten Impulsantwort wurde nach Kallinger et al. [85] mit einem linearen Prädiktorfehlerfilter der Länge $L_p = 20$ geglättet. Der Nachhall der geformten und nachgefilterten Impulsantwort beträgt $\mathrm{nPRQ} = 8.8\,\mathrm{dB}$, die spektrale Flachheit beträgt $\mathrm{SFM} = 0.53$.

Nachhall wurde auf $\mathrm{nPRQ} = 5.9\,\mathrm{dB}$ reduziert, die spektrale Flachheit wird mit $\mathrm{SFM} = 0.62$ quantifiziert.

Raumimpulsantwort aus dem Fraunhofer Gebäude Für die im Oldenburger Technologie- und Gründerzentrum gemessene Raumimpulsantwort gilt $\mathrm{nPRQ} = 23.5\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.69$. Für die Formung ohne Regularisierung ($\alpha = 0$) gilt $\mathrm{nPRQ} = 7.5\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.32$; die geformte globale Impulsantwort ist in Abbildung 6.7 dargestellt.

In Tabelle 6.3 ist der wahrnehmbare Nachhall (quantifiziert durch das nPRQ Maß aus [79]) und in Tabelle 6.4 die spektrale Flachheit für unterschiedliche Parameter des Optimierungsproblems dargestellt. In Abbildung 6.8 ist exemplarisch das sich ergebende Gesamtsystem bei gleichzeitiger Optimierung der mit $\alpha = 5$ gewichteten p-Norm ($p_f = 8$) des Betragsfrequenzgangs des Gesamtsystems dargestellt. Für das Gesamtsystem gilt hierbei $\mathrm{nPRQ} = 12.6\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.68$.

Im Wohnzimmer gemessene Raumimpulsantwort Für die im Wohn- und Esszimmer gemessene Raumimpulsantwort gilt $\mathrm{nPRQ} = 12.9\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.31$. Für

Setup \ α	TG Oldenburg, nPRQ in dB, p-Norm Regularisierung									
	0.1	0.25	0.5	0.75	1	2.5	5	10	20	40
$p_f = 8$, EQ	9.1	10.1	13.8	15.7	16.6	18.9	20.7	21.3	22.4	23.0
$p_f = 8$, GIR	8.1	8.8	9.1	9.5	9.4	10.4	12.6	14.8	16.3	17.0
$p_f = \infty$, EQ	14.3	16.0	17.5	18.6	19.6	22.4	23.5	23.5	23.5	23.5
$p_f = \infty$, GIR	9.1	12.9	14.6	15.2	15.6	19.4	22.3	22.6	23.0	23.2

Tabelle 6.3: Wahrnehmbarer Nachhall, quantifiziert in nPRQ [79]. Als Raumimpulsantwort wurde die im TG Oldenburg gemessene Impulsantwort verwendet. Für die nicht geformte Impulsantwort gilt $\mathrm{nPRQ} = 23.5\,\mathrm{dB}$, für die Formung ohne Regularisierung ergibt sich $\mathrm{nPRQ} = 7.5\,\mathrm{dB}$.

Setup \ α	TG Oldenburg, SFM, p-Norm Regularisierung									
	0.1	0.25	0.5	0.75	1	2.5	5	10	20	40
$p_f = 8$, EQ	0.40	0.46	0.65	0.69	0.70	0.70	0.69	0.69	0.69	0.69
$p_f = 8$, GIR	0.35	0.39	0.42	0.44	0.45	0.53	0.68	0.79	0.82	0.84
$p_f = \infty$, EQ	0.66	0.69	0.70	0.70	0.70	0.69	0.69	0.69	0.69	0.69
$p_f = \infty$, GIR	0.41	0.63	0.68	0.70	0.70	0.71	0.69	0.69	0.69	0.69

Tabelle 6.4: Spektrale Flachheit (SFM) nach der Formung. Als Raumimpulsantwort wurde die im TG Oldenburg gemessene Impulsantwort verwendet. Für die nicht geformte Impulsantwort gilt $\mathrm{SFM} = 0.69$, für die Formung ohne Regularisierung ergibt sich $\mathrm{SFM} = 0.32$.

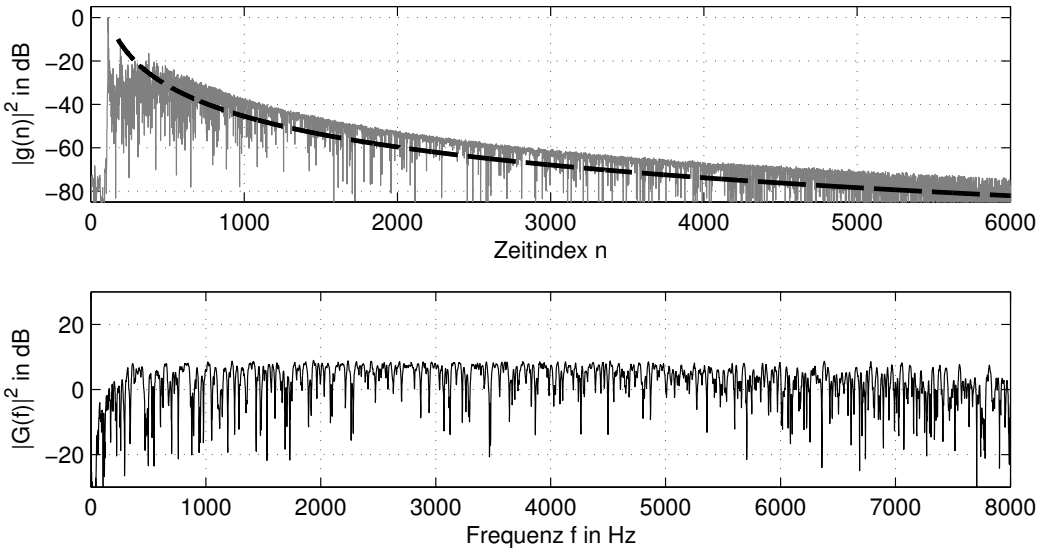

Abbildung 6.6: Ergebnis der Formung der mit dem Kunstkopf gemessenen Raumimpulsantwort mit Regularisierung durch die p-Norm ($\alpha = 5$, $p_f = 8$, GIR). Der Nachhall der geformten Impulsantwort beträgt $\mathrm{nPRQ} = 5.9\,\mathrm{dB}$, die spektrale Flachheit beträgt $\mathrm{SFM} = 0.62$.

die Formung ohne Regularisierung ($\alpha = 0$) gilt $\mathrm{nPRQ} = 0.0\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.03$; die geformte globale Impulsantwort ist in Abbildung 6.9 dargestellt.

In Tabelle 6.5 ist der wahrnehmbare Nachhall (quantifiziert durch das nPRQ Maß aus [79]) und in Tabelle 6.6 die spektrale Flachheit für unterschiedliche Parameter des Optimierungsproblems dargestellt. In Abbildung 6.10 ist exemplarisch das sich ergebende Gesamtsystem bei gleichzeitiger Optimierung der mit $\alpha = 20$ gewichteten p-Norm ($p_f = 8$) des Betragsfrequenzgangs des Gesamtsystems dargestellt. Für das Gesamtsystem gilt hierbei $\mathrm{nPRQ} = 7.1\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.43$.

Zusammenfassung - Regularisierung durch die p-Norm Die gleichzeitige Optimierung von Zeit- und Frequenzbereichsdarstellung ist einer Formung mit anschließender Entzerrung des Frequenzgangs vorzuziehen. Deutlich wird dies beim Vergleich der Abbildungen 6.5 und 6.6. Werden die in den Experimenten erzielten Werte für das nPRQ Maß sowie für das SFM in den Tabellen 6.1 bis 6.6 betrachtet, so ist generell festzustellen, dass eine Optimierung der Frequenzdarstellung des Gesamtsystems durch die p-Norm (mit $p_f = 8$) einer Bewertung durch die Maximumsnorm oder einer Optimierung der Frequenzdarstellung des zu entwerfenden Filters vorzuziehen ist.

Setup \ α	Wohnzimmer, nPRQ in dB, p-Norm Regularisierung									
	0.1	0.25	0.5	0.75	1	2.5	5	10	20	40
$p_f = 8$, EQ	0.0	5.1	7.3	7.6	8.6	10.1	11.4	11.6	12.4	13.4
$p_f = 8$, GIR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.1	8.8
$p_f = \infty$, EQ	5.6	7.8	9.1	9.8	11.4	12.3	12.9	12.9	12.9	12.9
$p_f = \infty$, GIR	0.0	0.0	0.0	4.0	5.2	8.5	10.5	11.9	11.9	11.8

Tabelle 6.5: Wahrnehmbarer Nachhall, quantifiziert in nPRQ [79]. Als Raumimpulsantwort wurde die im Wohn- und Esszimmer gemessene Impulsantwort verwendet. Für die nicht geformte Impulsantwort gilt $\mathrm{nPRQ} = 12.9\,\mathrm{dB}$, für die Formung ohne Regularisierung ergibt sich $\mathrm{nPRQ} = 0.0\,\mathrm{dB}$.

Setup \ α	Wohnzimmer, SFM, p-Norm Regularisierung									
	0.1	0.25	0.5	0.75	1	2.5	5	10	20	40
$p_f = 8$, EQ	0.06	0.28	0.32	0.32	0.33	0.32	0.31	0.31	0.30	0.30
$p_f = 8$, GIR	0.03	0.03	0.03	0.03	0.05	0.04	0.05	0.05	0.43	0.47
$p_f = \infty$, EQ	0.30	0.33	0.33	0.33	0.32	0.30	0.30	0.30	0.30	0.30
$p_f = \infty$, GIR	0.03	0.03	0.03	0.24	0.30	0.35	0.33	0.31	0.31	0.31

Tabelle 6.6: Spektrale Flachheit (SFM) nach der Formung. Als Raumimpulsantwort wurde die im Wohn- und Esszimmer gemessene Impulsantwort verwendet. Für die nicht geformte Impulsantwort gilt $\mathrm{SFM} = 0.31$, für die Formung ohne Regularisierung ergibt sich $\mathrm{SFM} = 0.03$.

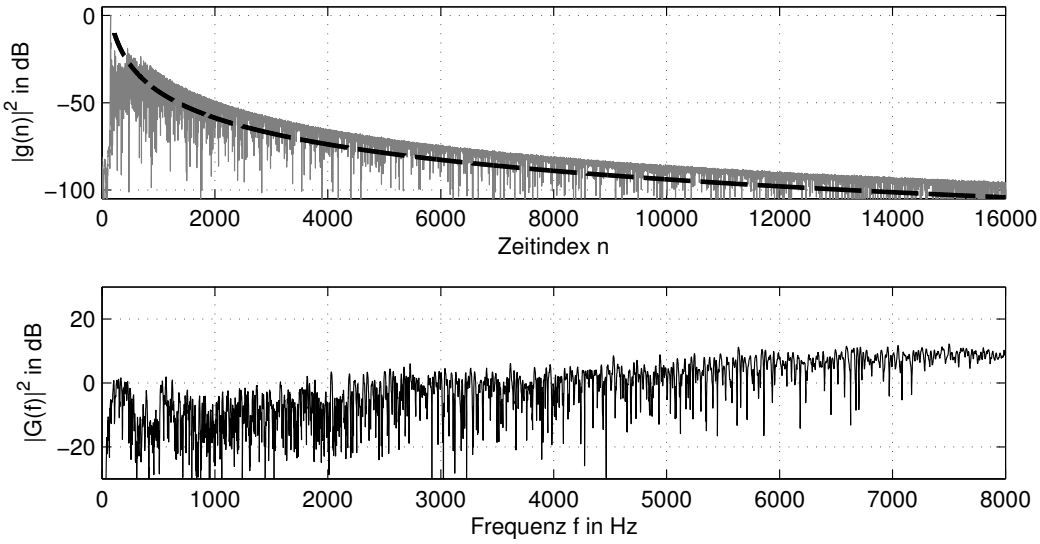

Abbildung 6.7: Ergebnis der Formung der im Technologie- und Gründerzentrum Oldenburg gemessenen Raumimpulsantwort ohne Regularisierung. Der Nachhall der geformten Impulsantwort beträgt $\mathrm{nPRQ} = 7.5\,\mathrm{dB}$, die spektrale Flachheit beträgt $\mathrm{SFM} = 0.32$.

6.2.1.2 Regularisierung auf Basis der spektralen Flachheit

In diesem Abschnitt werden die Ergebnisse bei gleichzeitiger Optimierung der p-Norm der Zeitbereichsdarstellung sowie der spektralen Flachheit nach (4.59) präsentiert. Hierzu wird für die drei in Kapitel 5 vorgestellten Raumimpulsantworten jeweils das in Gleichung (4.63) gegebene Optimierungsproblem gelöst.

Optimierung der spektralen Flachheit In einem ersten Experiment wurde das in Abschnitt 4.2.2.1 hergeleitete und in [75] publizierte generalisierte Maß zur Bewertung der spektralen Flachheit derart parametriert, dass direkt das aus der Literatur bekannte Maß zur Bewertung der spektralen Flachheit (SFM, [51]) maximiert wird. Diese Methode wurde in [74] publiziert.

Die Ergebnisse sind jeweils in tabellarischer Form dargestellt. In den Tabellen bezeichnet α den Gewichtsfaktor für den Regularisierungsterm aus (4.63). In Tabelle 6.7 sind die Ergebnisse für die mit dem Kunstkopf gemessene Raumimpulsantwort, in Tabelle 6.8 sind die Ergebnisse für die im TG Oldenburg gemessene Raumimpulsantwort und in Tabelle 6.9 sind die Ergebnisse für die im Wohn- und Esszimmer gemessene Raumimpulsantwort zusammengefasst.

	Kunstkopf, nPRQ in dB und SFM, gSFM Regularisierung										
α	0.1	0.25	0.5	0.75	1	2.5	5	7.5	10	20	40
nPRQ	2.2	2.5	6.0	6.4	6.7	8.2	9.7	10.9	11.7	14.0	15.8
SFM	0.32	0.40	0.65	0.70	0.74	0.84	0.90	0.92	0.94	0.96	0.98

Tabelle 6.7: Wahrnehmbarer Nachhall, quantifiziert in nPRQ [79], und spektrale Flachheit. Als Raumimpulsantwort wurde eine mit dem Kunstkopf gemessene Impulsantwort verwendet. Für die nicht geformte Impulsantwort gilt $\mathrm{nPRQ} = 11.6\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.58$, für die Formung ohne Regularisierung ergibt sich $\mathrm{nPRQ} = 2.1\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.24$.

	TG Oldenburg, nPRQ in dB und SFM, gSFM Regularisierung										
α	0.1	0.25	0.5	0.75	1	2.5	5	7.5	10	20	40
nPRQ	8.0	8.9	9.7	10.5	11.1	14.1	16.4	17.6	18.4	20.1	21.7
SFM	0.35	0.41	0.47	0.54	0.60	0.79	0.88	0.91	0.93	0.95	0.97

Tabelle 6.8: Wahrnehmbarer Nachhall, quantifiziert in nPRQ [79], und spektrale Flachheit. Als Raumimpulsantwort wurde die im TG Oldenburg gemessene Impulsantwort verwendet. Für die nicht geformte Impulsantwort gilt $\mathrm{nPRQ} = 23.5\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.69$, für die Formung ohne Regularisierung ergibt sich $\mathrm{nPRQ} = 7.5\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.32$.

	Wohnzimmer, nPRQ in dB und SFM, gSFM Regularisierung										
α	0.1	0.25	0.5	0.75	1	2.5	5	7.5	10	20	40
nPRQ	0.0	0.0	0.0	0.1	0.4	3.4	6.6	9.7	10.6	11.4	13.1
SFM	0.06	0.10	0.22	0.37	0.38	0.60	0.74	0.80	0.83	0.88	0.93

Tabelle 6.9: Wahrnehmbarer Nachhall, quantifiziert in nPRQ [79], und spektrale Flachheit. Als Raumimpulsantwort wurde die im Wohn- und Esszimmer gemessene Impulsantwort verwendet. Für die nicht geformte Impulsantwort gilt $\mathrm{nPRQ} = 12.9\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.31$, für die Formung ohne Regularisierung ergibt sich $\mathrm{nPRQ} = 0.0\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.03$.

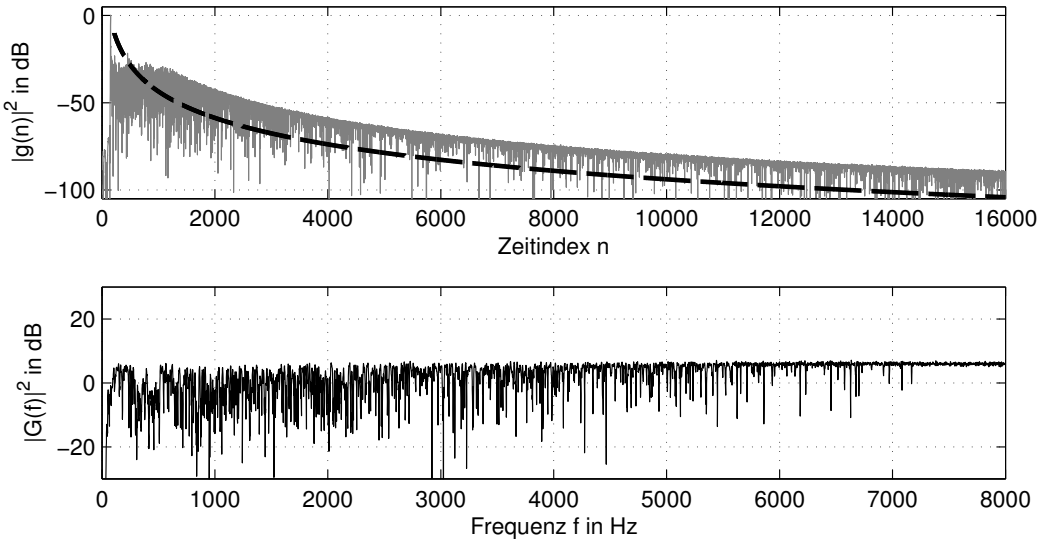

Abbildung 6.8: Ergebnis der Formung der im Technologie- und Gründerzentrum Oldenburg gemessenen Raumimpulsantwort mit Regularisierung durch die p-Norm ($\alpha = 5$, $p_f = 8$, GIR). Der Nachhall der geformten Impulsantwort beträgt $\mathrm{nPRQ} = 12.6\,\mathrm{dB}$, die spektrale Flachheit beträgt $\mathrm{SFM} = 0.68$.

Optimierung der spektralen Flachheit in einem Teilband Für das Experiment wurde das gSFM (4.59) derart parametriert, dass für die Berechnung des Maßes nicht die gesamte Bandbreite verwendet wird [74]. So können *don't care* Bereiche definiert werden, in denen der Betragsfrequenzgang beliebig geformt sein darf; für die Optimierung der Zeitbereichsdarstellung ist dieser Ansatz hinlänglich bekannt [94, 95, 170, 171].

Durch die Berechnung des Maßes zur Bewertung der spektralen Flachheit in einem Teilband kann die Optimierung auf die zu übertragenden Signale und/oder die zur Wiedergabe verwendete Hardware angepasst werden; beispielsweise ist die menschliche Sprache in einem Bereich von $f_L \approx 80\,\mathrm{Hz}$ bis $f_U \approx 12\,\mathrm{kHz}$ lokalisiert [160]. Werden lediglich Sprachsignale über den akustischen Kanal übertragen ist also die Flachheit des Betragsfrequenzgangs in diesem Bereich von besonderer Bedeutung für die wahrgenommene Qualität.

Für die Demonstration dieses Ansatzes wird die im Wohn- und Esszimmer gemessene Raumimpulsantwort (Abbildung 5.8) verwendet. Das gSFM wird derart parametriert, dass lediglich der Frequenzbereich zwischen $f_L = 20\,\mathrm{Hz}$ und $f_U = 20950\,\mathrm{Hz}$ zur Berechnung des Maßes verwendet wird. Die durch den zur Rekonstruktion verwendeten Tiefpassfilter entfernten hohen Frequenzen werden also nicht innerhalb des Filterentwurfes beachtet. Die Ergebnisse sind in tabellarischer Form in Tabelle 6.10 gegeben; das in

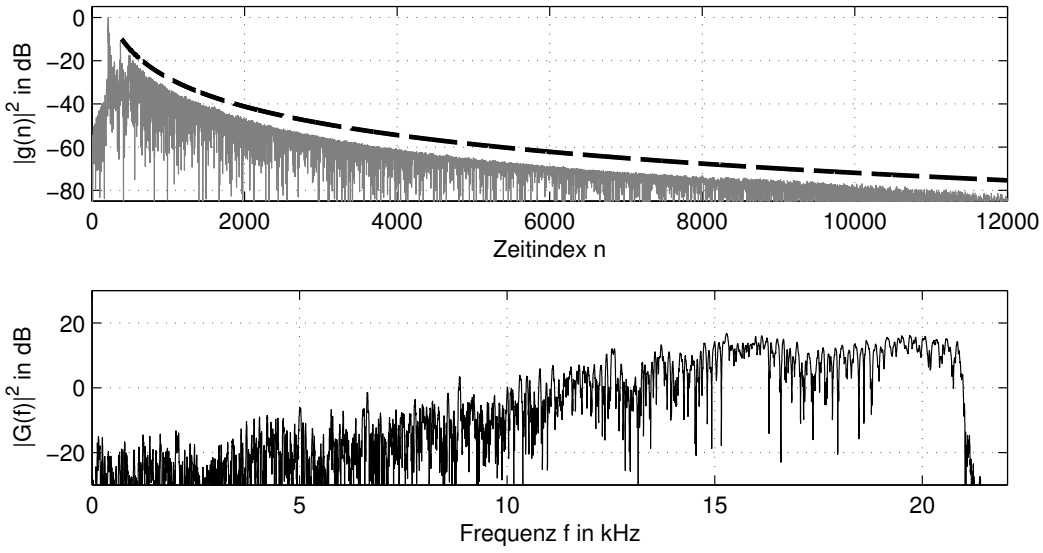

Abbildung 6.9: Ergebnis der Formung der im Wohn- und Esszimmer gemessenen Raumimpulsantwort ohne Regularisierung. Der Nachhall der geformten Impulsantwort beträgt $\mathrm{nPRQ} = 0.0\,\mathrm{dB}$, die spektrale Flachheit beträgt $\mathrm{SFM} = 0.03$.

der Tabelle angegebene SFM wurde der Parametrierung entsprechend lediglich für das Band von $f_L = 20\,\mathrm{Hz}$ bis $f_U = 20950\,\mathrm{Hz}$ berechnet.

In Abbildung 6.11 ist das Ergebnis der Formung für $\alpha = 2.5$ dargestellt; es wurde das SFM im Intervall $[20\,\mathrm{Hz} - 20.95\,\mathrm{kHz}]$ berücksichtigt. Der Nachhall der geformten Impulsantwort konnte auf $\mathrm{nPRQ} = 2.8\,\mathrm{dB}$ reduziert werden; die spektrale Flachheit im betrachteten Teilband beträgt $\mathrm{SFM} = 0.83$. Bei einer Formung ohne Regularisierung der Frequenzdarstellung (Abbildung 6.9) kann eine Reduktion des wahrnehmbaren Nachhalls auf $\mathrm{nPRQ} = 0.0\,\mathrm{dB}$ erreicht werden; wird die spektrale Flachheit im hier betrachteten Teilband berechnet, so ergibt sich $\mathrm{SFM} = 0.04$. Bei Regularisierung durch die p-Norm (siehe Abbildung 6.10) ergibt sich für das Gesamtsystem ein wahrnehmbarer Nachhall von $\mathrm{nPRQ} = 7.1\,\mathrm{dB}$ sowie eine spektrale Flachheit von $\mathrm{SFM} = 0.75$ im hier betrachteten Teilband.

Formung des Betragsfrequenzgangs durch Nachfilterung Für ein vergleichendes Experiment wurde für die mit dem Kunstkopf aufgenommene Raumimpulsantwort (Abbildung 5.5) ein Entzerrer der Länge $L_h = 3801$ entworfen. Zur Regularisierung wurde das gSFM derart parametriert, dass das Gesamtsystem einen flachen Betragsfrequenzgang aufweist. Der Gewichtsfaktor wurde als $\alpha = 0.5$ gewählt. Das Gesamtsystem

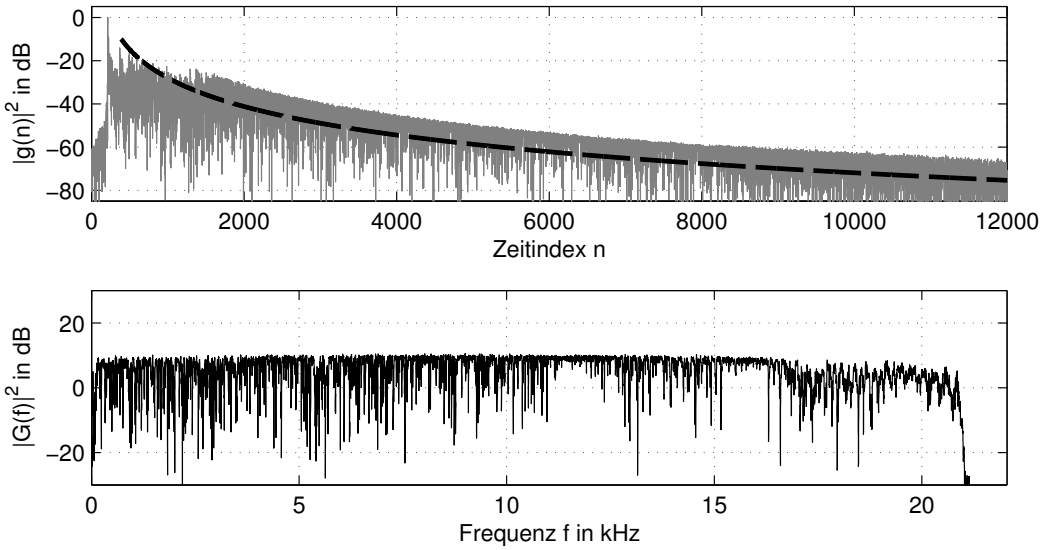

Abbildung 6.10: Ergebnis der Formung der im Wohn- und Esszimmer gemessenen Raumimpulsantwort mit Regularisierung durch die p-Norm ($\alpha = 20$, $p_f = 8$, GIR). Der Nachhall der geformten Impulsantwort beträgt $\mathrm{nPRQ} = 7.1\,\mathrm{dB}$, die spektrale Flachheit beträgt $\mathrm{SFM} = 0.43$.

der Länge $L_g = 7800$ wurde mit einem FIR Filter der Länge $L_p = 200$ nachgefiltert, um einen gewünschten Betragsfrequenzgang zu approximieren. Der zu approximierende Betragsfrequenzgang ist in Abbildung 6.12 dargestellt. Das Filter wurde mit der MATLAB Routine `firpm()` entworfen [138]. Die Parameter wurden entsprechend Abbildung 6.12 gewählt. Das sich so ergebende Gesamtsystem hat eine Länge von $L_g = 7999$ Koeffizienten und ist in Abbildung 6.13 dargestellt; das nPRQ Maß des Gesamtsystems beträgt $5.8\,\mathrm{dB}$.

Approximation eines beliebigen Betragsfrequenzgangs Das generalisierte Maß zur Bewertung der spektralen Flachheit kann derart parametriert werden, dass die globale Impulsantwort einen beliebigen Betragsfrequenzgang approximiert. Für die Experimente wurde die mit dem Kunstkopf gemessene Raumimpulsantwort aus Abbildung 5.5 verwendet.

Des Weiteren wurde ein fiktives Audiogramm simuliert, welches den frequenzabhängigen Hörverlust darstellt. Aus dem Audiogramm wurde die gewünschte Verstärkung jeder Frequenz abgeleitet. Die frequenzabhängige gewünschte Verstärkung ist in Abbildung 6.12 dargestellt. Abbildung 6.12 kann als *auf den Kopf gestelltes* Audiogramm verstanden werden. In Abbildung 6.14 sind die aus dem Audiogramm abgeleiteten

	Wohnzimmer, nPRQ in dB und SFM im Teilband [20 Hz − 20.95 kHz]										
α	0.1	0.25	0.5	0.75	1	2.5	5	7.5	10	20	40
nPRQ	0.0	0.0	0.0	0.1	0.5	2.8	4.8	6.0	6.9	8.8	10.3
SFM	0.08	0.15	0.32	0.51	0.62	0.83	0.90	0.92	0.94	0.96	0.97

Tabelle 6.10: Wahrnehmbarer Nachhall, quantifiziert in nPRQ [79], und spektrale Flachheit in einem Teilband. Als Raumimpulsantwort wurde die im Wohn- und Esszimmer gemessene Impulsantwort verwendet. Für den Filterentwurf wurde das gSFM (4.59) derart parametriert, dass lediglich das Teilband von $f_L = 20\,\mathrm{Hz}$ bis $f_U = 20950\,\mathrm{Hz}$ berücksichtigt wird. Für die nicht geformte Impulsantwort gilt $\mathrm{nPRQ} = 12.9\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.51$, für die Formung ohne Regularisierung ergibt sich $\mathrm{nPRQ} = 0.0\,\mathrm{dB}$ sowie $\mathrm{SFM} = 0.04$, wobei das SFM jeweils auf dem betrachteten Teilband berechnet wurde.

Gewichtsfaktoren γ_s in Abhängigkeit der Frequenz (notiert als γ_f) dargestellt (siehe Gleichung (4.61)).

In einem ersten Experiment wurden die Gewichtsfaktoren γ_s aus (4.59) derart gewählt, dass sie dem Reziproken der gewünschten Verstärkungsfaktoren entsprechen. Die weiteren Parameter waren $p_d = 20$, $p_u = 10$, $\alpha = 1$. Die Matrix $\boldsymbol{W}$ aus (4.69) wurde als Einheitsmatrix $\boldsymbol{I}$ gewählt; es wurde also keine Integration in Teilbänder vorgenommen. Das entworfene Filter hat eine Länge von $L_h = L_c = 4000$ Koeffizienten. Die so erzielten Ergebnisse sind in Abbildung 6.15 dargestellt. Der wahrnehmbare Nachhall wurde von $\mathrm{nPRQ} = 11.6\,\mathrm{dB}$ auf $\mathrm{nPRQ} = 4.1\,\mathrm{dB}$ reduziert.

In einem zweiten Experiment wurde die Matrix $\boldsymbol{W}$ derart gewählt, dass die K Koeffizienten der diskreten Fourier-Transformierten des globalen Systems in $S = 27$ Teilbänder integriert werden. Die Einträge w_{sk} der Matrix $\boldsymbol{W}$ wurden gehörgerecht einer Gammatone-Filterbank [64, 104, 159] entsprechend gewählt. Die Gewichtsfaktoren γ_s wurden entsprechend der Mittenfrequenzen des jeweiligen Gammatone-Bandes derart gewählt, dass sie dem aus dem Audiogramm abgeleiteten gewünschten Verstärkungsfaktor (Abbildung 6.14) entsprechen. Die derart gewichteten Gammatone-Bänder sind exemplarisch in Abbildung 6.16 auf einer logarithmischen Skala dargestellt.

Das sich ergebende Gesamtsystem ist in Abbildung 6.17 dargestellt; der wahrnehmbare Nachhall wurde auf $\mathrm{nPRQ} = 2.6\,\mathrm{dB}$ reduziert.

Zusammenfassung - Regularisierung duch das gSFM Die Optimierung des generalisierten Maßes zur Bewertung der spektralen Flachheit (4.59) erlaubt, beliebige Anforderungen an den Betragsfrequenzgang des Gesamtsystems zu stellen. Exemplarisch

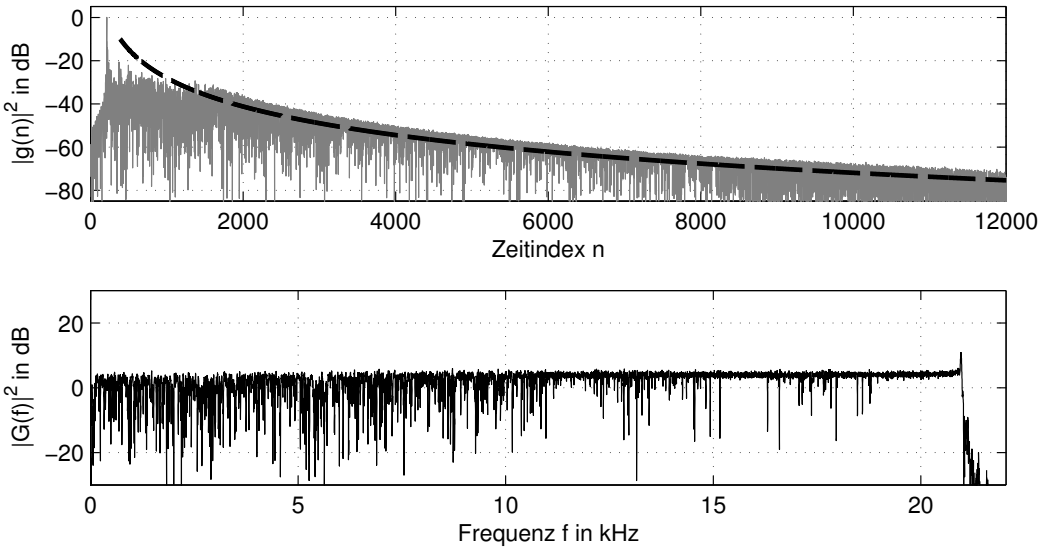

Abbildung 6.11: Ergebnis der Formung der im Wohn- und Esszimmer gemessenen Raumimpulsantwort bei Regularisierung durch das SFM im Teilband $[20\,\mathrm{Hz} - 20.95\,\mathrm{kHz}]$, gewichtet mit $\alpha = 2.5$. Der Nachhall der geformten Impulsantwort beträgt $\mathrm{nPRQ} = 2.8\,\mathrm{dB}$, die spektrale Flachheit im betrachteten Teilband beträgt $\mathrm{SFM} = 0.83$.

wurde dies gezeigt, indem ein fiktives Audiogramm definiert wurde. Das Gesamtsystem aus Raumimpulsantwort und Filter gleicht den frequenzabhängigen Hörverlust aus während gleichzeitig eine Reduktion des wahrnehmbaren Nachhalls erfolgt (siehe Abbildung 6.15). Auch in diesem Fall ist die gleichzeitige Optimierung von Zeit- und Frequenzbereichsdarstellung einer zweistufigen Methode vorzuziehen; deutlich wird dies beim Vergleich der Abbildungen 6.15 und 6.13.

Das gSFM erlaubt durch entsprechende Parametrierung, die Eigenschaften betrachteten Signale und/oder der Wiedergabehardware zu berücksichtigen. Exemplarisch wurde hierfür der Effekt des Rekonstruktionstiefpassfilters eines D/A Wandlers während der Optimierung berücksichtigt (siehe Abbildung 6.11).

Das Konzept der Integration in Teilbänder erlaubt eine weitere Reduktion des wahrnehmbaren Nachhalls, was beim Vergleich der Abbildungen 6.15 (keine Integration) und 6.17 (Integration in $S = 27$ gehörrichtige Teilbänder durch eine Gammatone-Filterbank [64, 104]) deutlich wird; die Anforderungen an den gewünschten Betragsfrequenzgang werden durch die Integration abgeschwächt. Durch dieses Konzept können psychoakustische Eigenschaften des menschlichen Hörsystems ausgenutzt werden.

Bei einem Vergleich der Ergebnisse, die durch Optimierung des gSFM Maßes erzielt wurden mit den Ergebnissen der auf der p-Norm basierenden Regularisierung

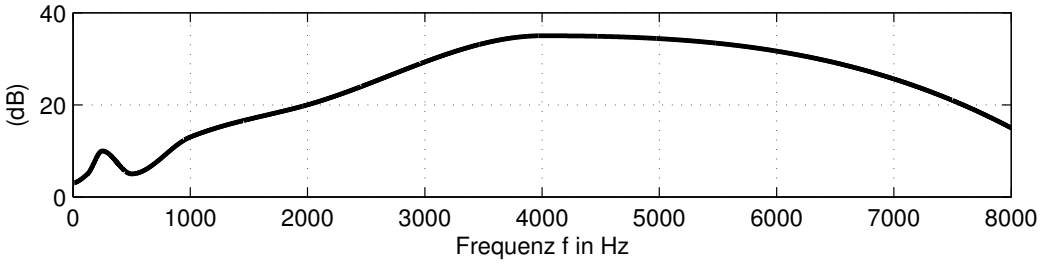

Abbildung 6.12: Aus einem fiktiven Audiogramm abgeleitete gewünschte frequenzabhängige Verstärkung. Dieses Bild kann als gespiegeltes Audiogramm angesehen werden.

aus Abschnitt 6.2.1.1 ist zu erkennen, dass die Regularisierung durch das gSFM im Allgemeinen zu besseren Ergebnissen (quantifiziert durch nPRQ [79] und SFM [105]) führt.

6.2.2 Räumliche Robustheit

In diesem Abschnitt werden die Ergebnisse zur räumlichen Robustheit präsentiert. Als Datengrundlage wurden die mit dem Kunstkopf gemessenen Raumimpulsantworten verwendet. Es wurden die am linken Ohr des Kunstkopfes auf Position P_1 (siehe Abbildung 5.3) gemessenen Raumimpulsantworten verwendet. Für die Reproduktion wurden alle $N = 4$ Lautsprecher verwendet. In allen durchgeführten Experimenten haben die Vorfilter eine Länge von $L_h = 5000$ Zeitkoeffizienten; die Raumimpulsantworten haben eine Länge von $L_c = 4000$ Zeitkoeffizienten. Als Gewichtsfenster für die Zeitbereichsdarstellung wurden die in [120] empfohlenen Gewichtsfenster verwendet, sowie $p_d = 20$ und $p_u = 10$ als Parameter für die p-Norm.

Für die robusten Entwürfe auf Grundlage der im Rahmen dieser Arbeit entwickelten Verfahren wurden lediglich die Raumimpulsantworten von den Lautsprechern zur zentralen Referenzposition ($x = y = z = 0\,\mathrm{cm}$, in Abbildung 5.6) verwendet.

Die Evaluierung der Ergebnisse erfolgt, indem die globalen Impulsantworten an den 40 Testpositionen (siehe Abbildung 5.6) berechnet wurden. Für diese 40 akustischen Kanäle wurde jeweils das nPRQ Maß (6.1) und das SFM (4.57) berechnet; als Ergebnis werden die durchschnittlichen Werte der beiden Maße über die Testpositionen angegeben. Gemittelt über die an den 40 Testpositionen gemessenen Raumimpulsantworten beträgt das nPRQ Maß $9.9\,\mathrm{dB}$ und das SFM beträgt 0.63.

Wird ein nicht robuster Entwurf der Vorfilter zur Formung für die zentrale Referenzposition ($x = y = z = 0\,\mathrm{cm}$, in Abbildung 5.6) durch Optimierung von (4.47) mit $\alpha = 40$

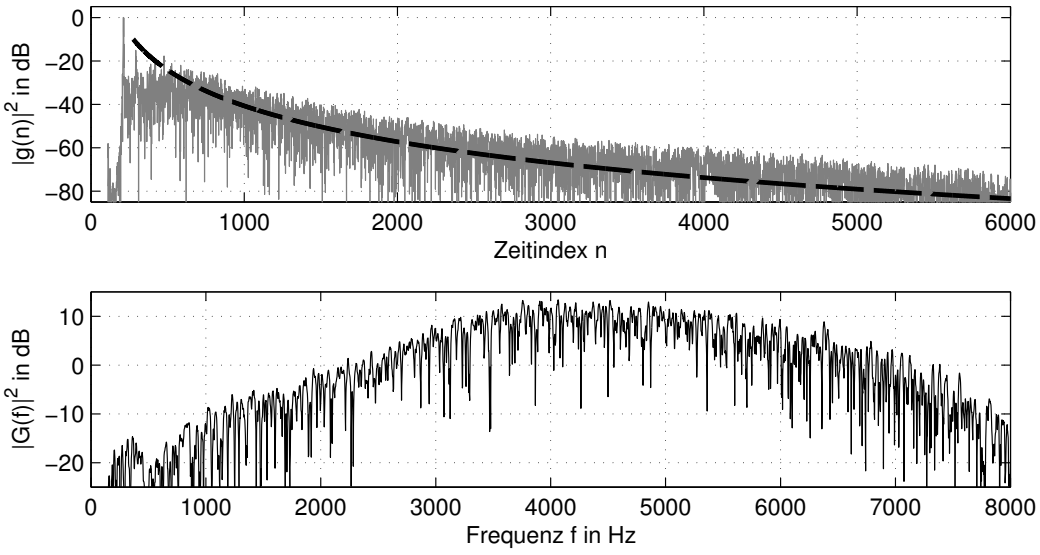

Abbildung 6.13: Die mit dem Kunstkopf gemessene Raumimpulsantwort wurde derart geformt, dass sich ein flacher Betragsfrquenzgang ergibt. In einem zweiten Schritt wurde ein FIR Filter mit Hilfe der `firpm()` Funktion aus MATLAB entworfen, welches den gewünschten Betragsfrequenzgang herstellt [138]. Der wahrnehmbare Nachhall des Gesamtsystems beträgt $\mathrm{nPRQ} = 5.8\,\mathrm{dB}$.

durchgeführt, so ergeben sich gemittelt über die 40 Testpositionen Werte von $10.5\,\mathrm{dB}$ für das nPRQ Maß und 0.64 für das Maß zur Bewertung der spektralen Flachheit.

In Abbildung 6.18 ist das Ergebnis der Formung der Raumimpulsantwort an der Referenzposition dargestellt; es wurde ein nicht-robuster Filterentwurf nach (4.48) mit $\alpha = 40$ (Optimierung des Betragsfrequenzgangs des Gesamtsystems) durchgeführt. In Abbildung 6.19 ist die globale Impulsantwort an einer der Testpositionen beim nicht-robusten Filterentwurf dargestellt.

6.2.2.1 Räumliche Robustheit durch Abtastung des Hörvolumens

Für den robusten Entwurf auf Grundlage des räumlichen Abtasttheorems für Raumimpulsantworten [1, 114] wurden die globalen Impulsantworten an den $R = 27$ Referenzpositionen (siehe Abbildung 5.6) gemeinsam geformt. Hierzu wird das durch (4.48) gegebene Optimierungsproblem gelöst.

Im erarbeiteten Formelwerk (siehe Abschnitt 4.1.2) wird der Filterentwurf mit $N = 4$ Lautsprechern, $R = 27$ Instanzen der Raumimpulsantwort und $M = 1$ Mikrofonposition durchgeführt. Die Ergebnisse sind in tabellarischer Form in Tabelle 6.11 zusammenge-

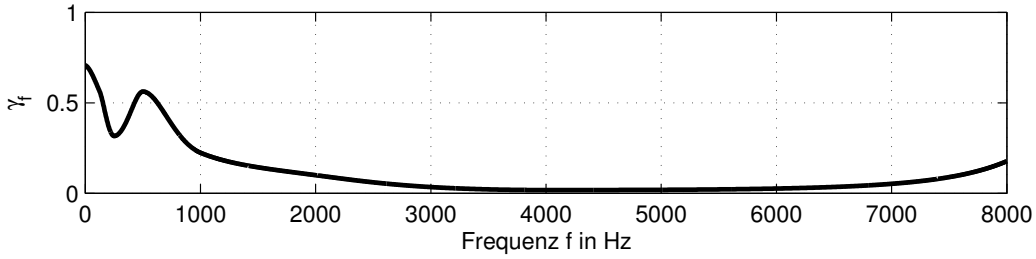

Abbildung 6.14: Aus einem fiktiven Audiogramm abgeleitete Gewichtsfaktoren. Das Gesamtsystem soll den in Abbildung 6.12 dargestellten gewünschten Betragsfrequenzgang approximieren.

fasst. Für alle Experimente wurde $p_f = 8$ gewählt. Die Abkürzung EQ beziehungsweise GIR notiert, ob die p-Norm des Betragsfrequenzgangs des Entzerrers beziehungsweise der globalen Impulsantwort optimiert wurde (siehe Gleichung (4.36)), und α ist der Gewichtsfaktor für den Regularisierungsterm. Sofern nicht explizit anders angegeben, wird der auf der p-Norm basierende Regularisierungsterm mit $p_f = 8$ parametriert.

Robuste Raumhallkompensation durch den Multipositions-Ansatz		
Konfiguration	nPRQ in dB	SFM
Ungeformt	9.9	0.63
nicht robust, $\alpha = 40$	10.5	0.64
robust, $\alpha = 0$	0.2	0.30
robust, $\alpha = 1$, EQ	0.9	0.54
robust, $\alpha = 2$, EQ	3.2	0.56
robust, $\alpha = 5$, GIR	0.8	0.62
robust, $\alpha = 10$, GIR	1.6	0.77

Tabelle 6.11: Ergebnisse für den robusten Entwurf auf Grundlage des räumlichen Abtasttheorems von Raumimpulsantworten. Die Maße wurden jeweils über die 40 Testpositionen gemittelt.

In Abbildung 6.20 ist die globale Impulsantwort an einer der Testpositionen beim robusten Filterentwurf auf Grundlage des räumlichen Abtasttheorems für Raumimpulsantworten [1, 114] dargestellt; hierbei wurde neben der Zeitbereichsdarstellung die p-Norm der Betragsfrequenzgänge der Gesamtsysteme optimiert und mit $\alpha = 5$ gewichtet.

Zusammenfassung - Mehrpositions-Ansatz In Tabelle 6.11 ist zu erkennen, dass der nicht-robuste Filterentwurf im Falle einer räumlichen Fehlplatzierung (Abbildung 6.19) zu keiner Reduktion des wahrnehmbaren Nachhalls führt. Erfolgt der

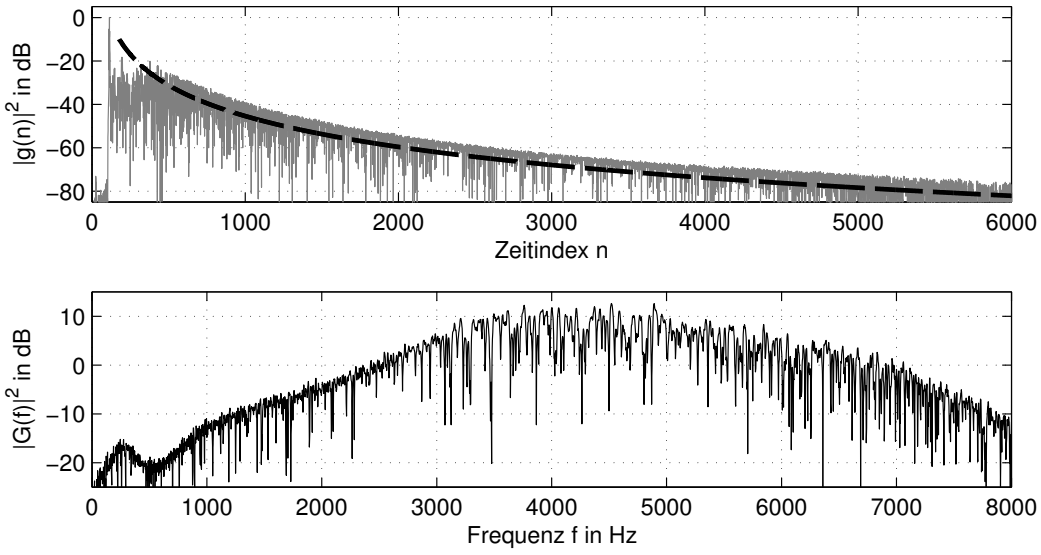

Abbildung 6.15: Ergebnis der Formung der mit dem Kunstkopf gemessenen Raumimpulsantwort durch Optimierung von (4.63). Es wurde keine Integration in Teilbänder vorgenommen. Die Gewichtsfaktoren γ_s aus Gleichung (4.59) wurden aus dem fiktiven Audiogramm abgeleitet (siehe Abbildung 6.14). Der hörbare Nachhall wurde auf $\mathrm{nPRQ} = 4.1\,\mathrm{dB}$ reduziert.

Filterentwurf für alle Referenzpositionen gleichzeitig, so wird eine Reduktion des wahrnehmbaren Nachhalls innerhalb des gesamten Hörvolumens (exemplarisch in Abbildung 6.20 dargestellt) erreicht. Auch für diesen Ansatz ist die gleichzeitige Optimierung von Zeit- und Frequenzbereichsdarstellung des Gesamtsystems einer Optimierung der Zeitbereichsdarstellung des Gesamtsystems bei gleichzeitiger Optimierung der Frequenzbereichsdarstellung der Filter vorzuziehen (Zeilen mit EQ beziehungsweise GIR in Tabelle 6.11).

6.2.2.2 Räumliche Robustheit durch statistische Modellierung

In diesem Abschnitt werden die Ergebnisse für den robusten Filterentwurf durch statistische Fehlermodellierung (Gleichung (4.112)) präsentiert. Die Filter zur Formung wurden hierbei für die zentrale Referenzposition ($x = y = z = 0\,\mathrm{cm}$, in Abbildung 5.6) entworfen.

In Tabelle 6.12 sind tabellarisch die über die 40 Testpositionen gemittelten Werte für das nPRQ Maß sowie für das SFM für exemplarische Parameter aufgetragen. Hierbei bezeichnet α den Gewichtsfaktor für den auf der p-Norm basierenden Regularisierungsterm ($p_f = 8$, Optimierung des Betragsfrequenzgangs des Gesamtsystems), β

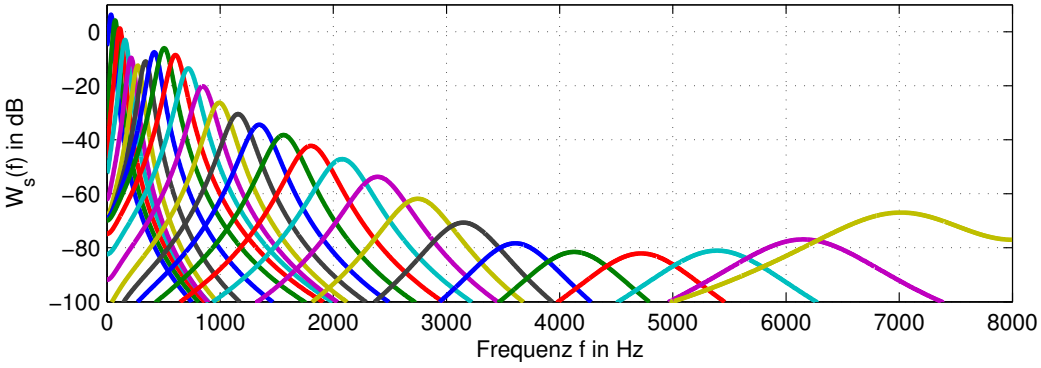

Abbildung 6.16: Dem gewünschten Betragsfrequenzgang entsprechend gewichtete Gammetone-Filterbank auf einer logarithmischen Skala. Die Abbildung stellt die $S = 27$ Zeilen der Matrix $\mathbf{\Psi}$ aus Gleichung (4.72) dar, welche die Integration in Teilbänder sowie die Gewichtung der Teilbänder abbildet.

bezeichnet den Gewichtsfaktor für den statistischen Störterm, D bezeichnet die für die Fehlermodellierung angenommene Fehlplatzierung.

Robuste Raumhallkompensation durch statistische Fehlermodellierung		
Parameter	nPRQ in dB	SFM
Ungeformt	9.9	0.63
nicht robust, $\alpha = 40$	10.5	0.64
robust, $D = 0.5\,\mathrm{cm}$, $\alpha = 1$, $\beta = 2{\cdot}10^{-4}$	5.0	0.62
robust, $D = 0.5\,\mathrm{cm}$, $\alpha = 1$, $\beta = 5{\cdot}10^{-4}$	4.5	0.64
robust, $D = 1\,\mathrm{cm}$, $\alpha = 1$, $\beta = 2{\cdot}10^{-4}$	4.7	0.60
robust, $D = 1\,\mathrm{cm}$, $\alpha = 1$, $\beta = 5{\cdot}10^{-4}$	4.1	0.61
robust, $D = 2\,\mathrm{cm}$, $\alpha = 1$, $\beta = 3{\cdot}10^{-4}$	3.9	0.60
robust, $D = 2\,\mathrm{cm}$, $\alpha = 1$, $\beta = 5{\cdot}10^{-4}$	3.3	0.52

Tabelle 6.12: Ergebnisse für den robusten Filterentwurf durch Optimierung von (4.107). α ist der Gewichtsfaktor für den auf der p-Norm des Betragsfrequenzgangs des Gesamtsystems basierenden Regularisierungsterm, β ist der Gewichtsfaktor für den statistischen Fehlerterm und D ist die angenommene durchschnittliche Fehlplatzierung zur Modellierung des Fehlerterms. Die Maße wurden jeweils über die 40 Testpositionen gemittelt.

In Abbildung 6.21 ist eine der globalen Impulsantworten im Falle von räumlicher Fehlplatzierung dargestellt; die Parameter zum Filterentwurf waren $D = 2\,\mathrm{cm}$, $\alpha = 1$ sowie $\beta = 3{\cdot}10^{-4}$.

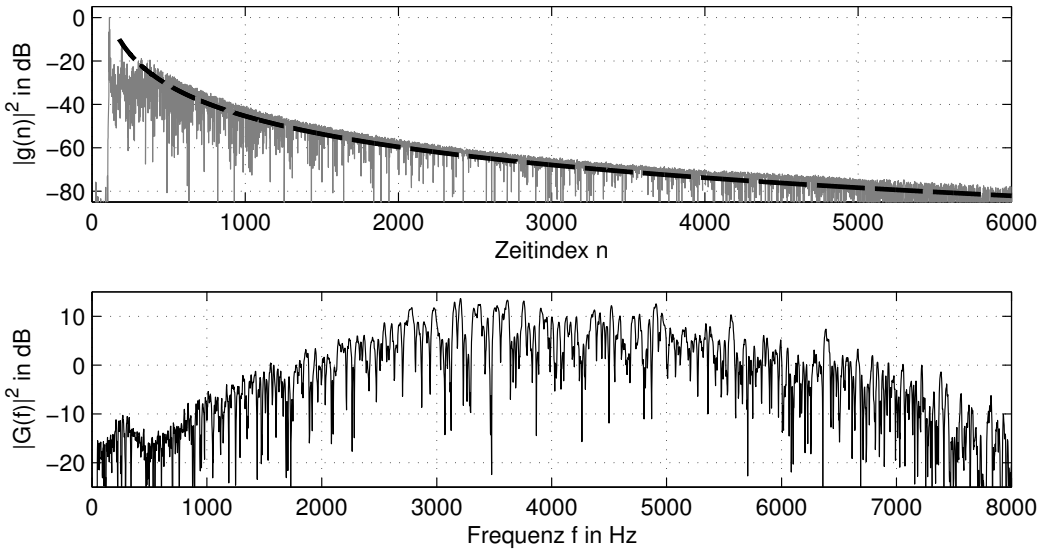

Abbildung 6.17: Ergebnis der Formung der mit dem Kunstkopf gemessenen Raumimpulsantwort durch Optimierung von (4.63). Es wurde eine Integration in $S = 27$ Teilbänder durch eine Gammatone-Filterbank vorgenommen (Abbildung 6.16). Die Gewichtsfaktoren γ_s aus Gleichung (4.59) wurden aus dem fiktiven Audiogramm entsprechend der Mittenfrequenzen der Gammatone-Bänder abgeleitet (siehe Abbildung 6.14). Der hörbare Nachhall wurde auf $\mathrm{nPRQ} = 2.6\,\mathrm{dB}$ reduziert.

Beschleunigung durch optimierte Berechnungen Die Berechnung der für den robusten Filterentwurf durch Optimierung von (4.112) nötigen Matrizen $\boldsymbol{M}_{m\ell}$ nach der in (4.98) gegebenen Berechnungsvorschrift ist sehr zeitaufwendig. In Abschnitt 4.3.1.5 wurde eine effiziente Berechnungsvorschrift für die benötigten Matrizen vorgestellt.

In Tabelle 6.13 ist tabellarisch die Dauer für die Berechnung einer einzelnen Matrix $\boldsymbol{M}_{m\ell}$ für unterschiedliche Längen L_c der Raumimpulsantwort und unterschiedliche Längen L_h des zu entwerfenden Filters aufgetragen. Die Messungen wurden auf einem PC mit einem einkernigen Prozessor mit $3\,\mathrm{Ghz}$ Taktfrequenz durchgeführt. In der Tabelle 6.13 bezeichnet A.1 die Berechnungsvorschrift nach (4.98); A.2 bezeichnet die optimierte Berechnungsvorschrift nach (4.119). Die Zeitangaben sind in Sekunden gegeben.

Zusammenfassung - Statistische Fehlermodellierung Wie in Tabelle 6.12 zu erkennen ist kann durch die statistische Modellierung des Fehlerterms eine räumlich robuste Nachhallreduktion innerhalb des gesamten Abhörvolumens erreicht werden. Im

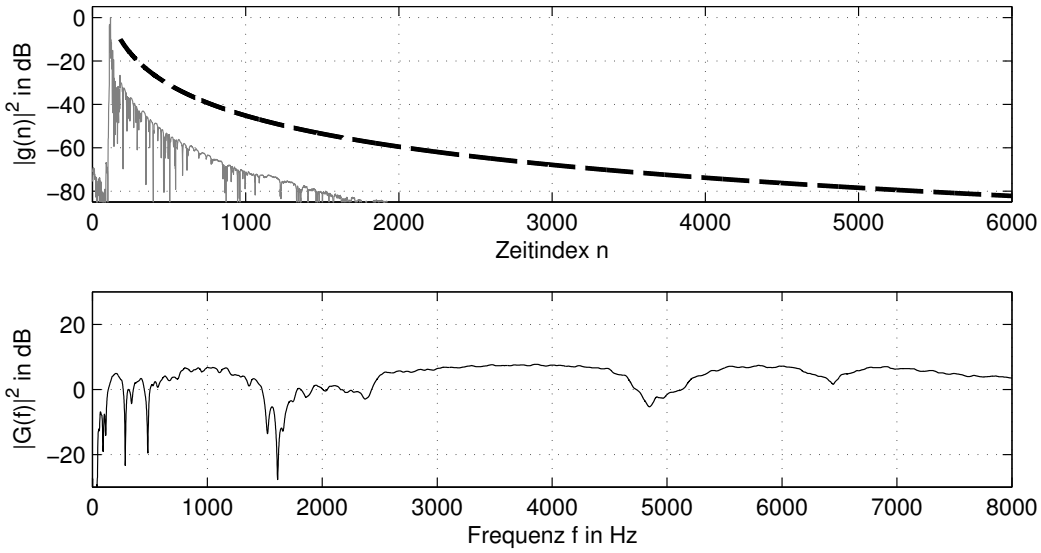

Abbildung 6.18: Globale Impulsantwort an der Referenzposition beim nicht-robusten Filterentwurf nach (4.48) mit $\alpha = 40$.

Beschleunigung der Vorberechnungen				
L_c	L_h	A.1	A.2	Beschleunigungsfaktor
1000	1000	926 s	4.3 s	**215**
2000	2000	14139 s	24.4 s	**580**
4000	5000	367112 s	226.3 s	**1622**

Tabelle 6.13: Tabellarische Gegenüberstellung der für die Berechnung der Matrizen $\boldsymbol{M}_{m\ell}$ notwendige Zeitaufwand. A.1 bezeichnet die Berechnung nach (4.98), A.2 bezeichnet die effiziente Berechnung nach (4.119). Alle Zeitangaben sind in Sekunden gegeben; die Messungen wurden auf einem PC mit 3 Ghz Taktfrequenz durchgeführt.

Vergleich zu dem Mehrpositions-Ansatz aus Abschnitt 6.2.2.1 kann hierbei auf eine exzessive Messung von Raumimpulsantworten verzichtet werden; der Filterentwurf erfolgt lediglich für die Referenzposition.

Die erzielte Nachhallreduktion innerhalb des Abhörvolumens bei statistischer Modellierung des Fehlerterms erreicht nicht die Güte der Nachhallreduktion, die durch eine Abtastung des Abhörvolumens erreicht werden kann. Dies wird beim Vergleich der in den Tabellen 6.11 beziehungsweise 6.12 dargestellten Ergebnisse deutlich.

Die effiziente Berechnung der benötigten Matrizen ist essentiell für die praktische Nutzbarmachung des Verfahrens. Hierbei wurde durch die effiziente Berechnungsvor-

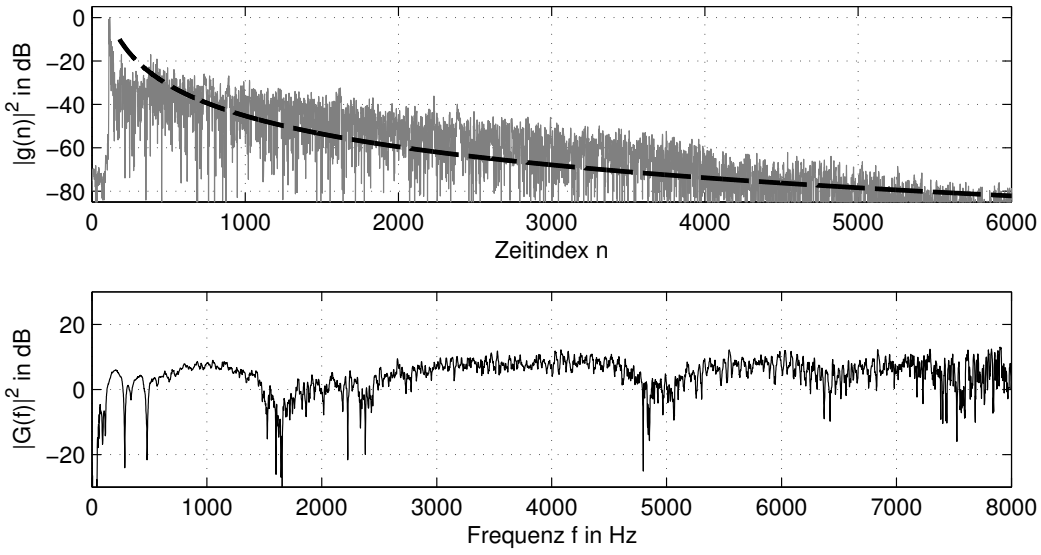

Abbildung 6.19: Globale Impulsantwort an einer der Testpositionen bei einem nicht-robusten Filterentwurf nach (4.48) mit $\alpha = 40$.

schrift aus Abschnitt 4.3.1.5 eine Beschleunigung um mehr als den Faktor 1000 erreicht (siehe Tabelle 6.13).

6.2.2.3 Räumliche Robustheit durch perturbierte Raumimpulsantworten

In diesem Abschnitt werden die Ergebnisse der räumlich robusten Impulsformung nach dem in Abschnitt 4.3.2 vorgestellten Verfahren präsentiert. In einem ersten Experiment werden die theoretisch geforderten (Abschnitt 4.3.2.2 und 4.3.2.3) Eigenschaften des Perturbationsterms mit den tatsächlichen Eigenschaften der künstlich erzeugten Perturbationen verglichen. Die künstlich erzeugten Perturbationen wurden nach dem in Abschnitt 4.3.2.4 präsentierten Verfahren generiert.

In einer zweiten Reihe von Experimenten wird auf Grundlage der zur Referenzposition gemessenen Raumimpulsantwort jeweils eine Schar von mehreren künstlich perturbierten Raumimpulsantworten generiert. Die Filter zur Impulsformung werden dann für diese Schar an Impulsantworten entworfen.

Künstlich generierte Perturbationen In Abbildung 6.22 sind exemplarisch die gewünschten Eigenschaften des Perturbationsterms bei einer angenommenen Fehlplatzierung von $D = 2\,\mathrm{cm}$ im Zeitbereich (links, nach Gleichung (4.129)) sowie im Frequenzbereich (rechts, nach Gleichung (4.130)) dargestellt. Als Referenz wurde die in Abbildung 5.5 dargestellte Raumimpulsantwort verwendet. In Abbildung 6.23 ist das

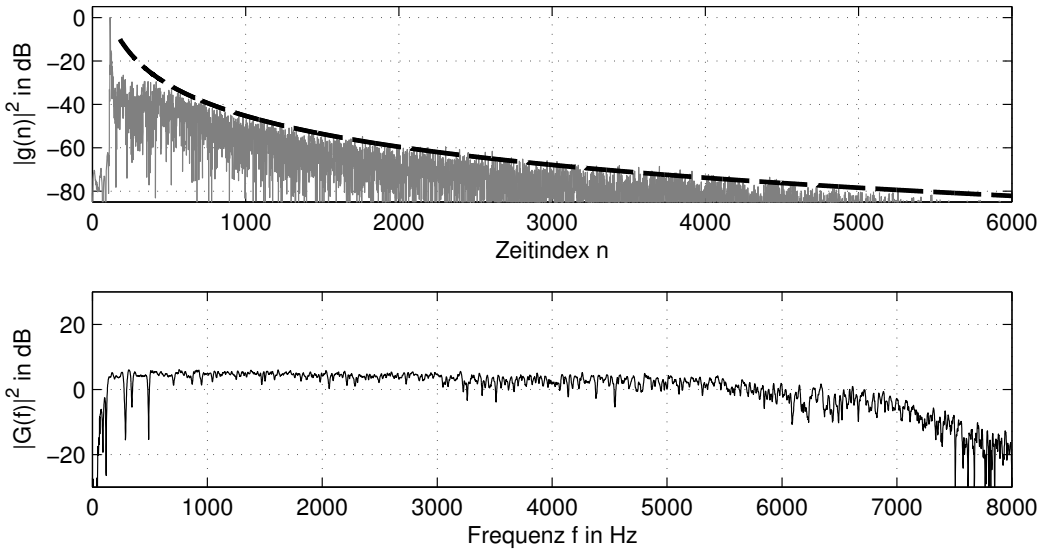

Abbildung 6.20: Globale Impulsantwort an einer der Testpositionen beim robusten Filterentwurf nach (4.48) unter Berücksichtigung des räumlichen Abtasttheorems für Raumimpulsantworten [1, 114]; der auf der p-Norm basierende Regularisierungsterm wurde mit $\alpha = 5$ gewichtet.

durchschnittliche Abklingverhalten des Perturbationsterms sowie die durchschnittliche frequenzabhängige Energie des Perturbationsterms dargestellt; angenommen wurde eine Fehlplatzierung von $D = 2\,\mathrm{cm}$. Für die Abbildung wurden 5000 Instanzen des Perturbationsterms $p(n)$ nach dem in Abschnitt 4.3.2.4 vorgestellten Verfahren generiert. Es wurde jeweils das Betragsquadrat der Zeitkoeffizienten sowie der quadrierte Betragsfrequenzgang berechnet und über die 5000 Instanzen gemittelt. Es ist deutlich zu erkennen, dass die gewünschten Eigenschaften aus Abbildung 6.22 approximiert werden.

Untersuchung zur räumlichen Robustheit In Tabelle 6.14 sind die Ergebnisse für den robusten Filterentwurf auf Basis der künstlich perturbierten Raumimpulsantwort dargestellt. In der Tabelle bezeichnet α den Gewichtsfaktor für den auf der p-Norm ($p_f = 8$) basierenden Regularisierungsterm zur Optimierung der Betragsfrequenzgänge der Gesamtsysteme, D ist die angenommene Fehlplatzierung, $\widehat{R}$ ist die Anzahl der künstlich erzeugten Instanzen der akustischen Kanäle und M_{dB} bezeichnet den normalisierten Systemfehler (4.138).

Die Filter zur Impulsformung werden gemeinsam für die gemessene Raumimpulsantwort und die $\widehat{R}$ künstlich perturbierten Instanzen der Raumimpulsantwort entworfen; die

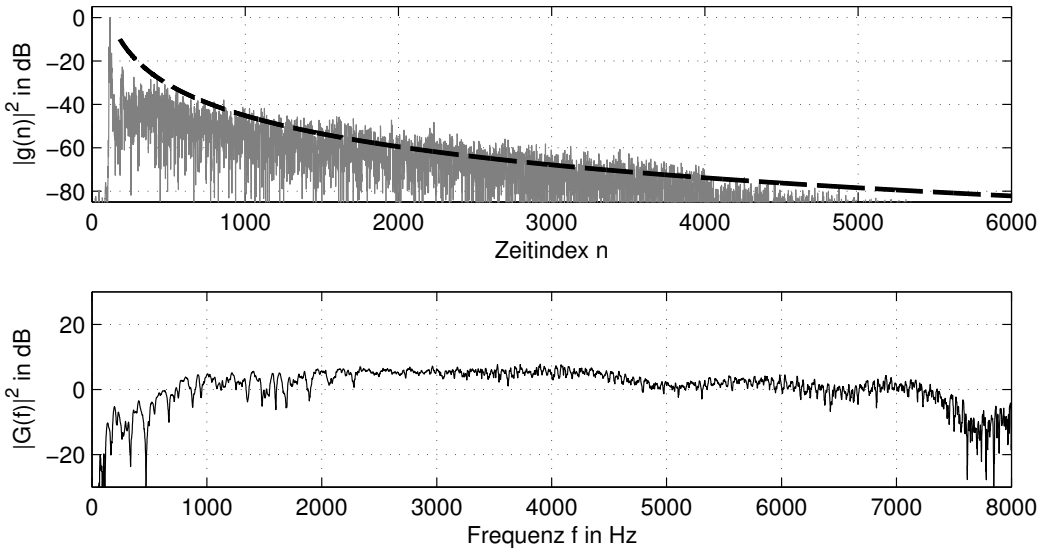

Abbildung 6.21: Globale Impulsantwort an einer der Testpositionen beim robusten Filterentwurf nach (4.107). Die zum Filterentwurf verwendeten Parameter waren $D = 2\,\mathrm{cm}$, $\alpha = 1$ (Optimierung des Betragsfrequenzgangs des Gesamtsystems) und $\beta = 3{\cdot}10^{-4}$.

Entzerrer werden also für $R = \widehat{R} + 1$ Instanzen der akustischen Kanäle von den vier Lautsprechern zur Referenzposition entworfen (vergleiche Abbildung 4.2).

In Abbildung 6.24 ist das Ergebnis der robusten Impulsformung bei räumlicher Fehlplatzierung dargestellt. Zum Filterentwurf wurden $\widehat{R} = 29$ perturbierte Instanzen der von den vier Lautsprechern zur Referenzposition gemessenen Raumimpulsantworten verwendet. Die Parameter für die Generierung der Perturbationen waren $D = 1\,\mathrm{cm}$, $M_{\mathrm{dB}} = 20\,\mathrm{dB}$; der auf der p-Norm basierende Regularisierungsterm zur Optimierung der Betragsfrequenzgänge der Gesamtsysteme wurde mit $\alpha = 2$ gewichtet.

Zusammenfassung - Perturbationsbasierter Ansatz Die Verwendung von perturbierten Instanzen der Raumimpulsantwort zur Referenzposition für den Filterentwurf erlaubt eine Nachhallreduktion innerhalb des gesamten Abhörvolumens. Die Nachhallreduktion durch dieses Verfahren erreicht nicht die Güte, die durch eine Abtastung des gesamten Abhörvolumens erreicht werden kann (vergleiche Tabelle 6.11 und 6.14). Im Vergleich zum robusten Filterentwurf durch eine statistische Modellierung des Fehlers erreicht der perturbationsbasierte Ansatz vergleichbare Werte bezüglich nPRQ und SFM im Falle einer räumlichen Fehlplatzierung (siehe Tabelle 6.12 und 6.14).

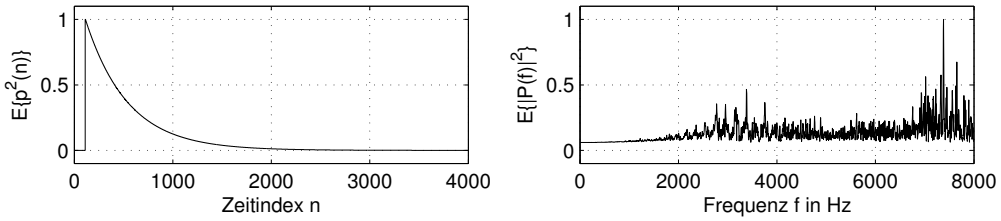

Abbildung 6.22: Geforderte Eigenschaften der Perturbationen im Zeit- (links) und Frequenzbereich (rechts). Die gewünschten Eigenschaften ergeben sich aus Gleichung (4.129) und Gleichung (4.130) für eine angenommene Fehlplatzierung von $D = 2\,\mathrm{cm}$ für die in Abbildung 5.5 dargestellte Raumimpulsantwort.

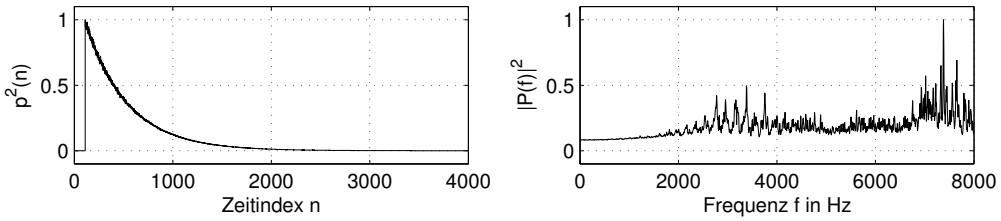

Abbildung 6.23: Durchschnittliches Abklingverhalten der Perturbationen im Zeitbereich (links). Durchschnittliche Energie des Perturbationsterms über die Frequenzen (rechts). Die Grafiken wurden über 5000 Instanzen für eine angenommene Fehlplatzierung von $D = 2\,\mathrm{cm}$ gemittelt.

Für den perturbationsbasierten Ansatz müssen $N \cdot R$ Raumimpulsantworten der Länge L_c im Speicher vorgehalten werden; für den statistischen Ansatz aus Abschnitt 4.3.1 müssen hingegen N Matrizen der Dimension $L_h \times L_h$ im Speicher vorgehalten werden. Diese Reduzierung der zur Berechnung benötigten Daten ist ein wichtiger Schritt zur praktischen Nutzbarmachung der robusten Raumhallkompensation in eingebetteten Systemen mit beschränktem Speicher.

6.2.3 Übersprechkompensation

In diesem Abschnitt werden die Ergebnisse zur gemeinsamen Übersprechkompensation und Enthallung präsentiert; die Ergebnisse wurden in [79] publiziert.

Für die gemeinsame Übersprech- und Raumhallkompensation wurde der in Abschnitt 5.3 vorgestellte Datensatz verwendet. Es wurden Experimente für zwei Lautsprecher und zwei Abhörpositionen, vier Lautsprecher und zwei Abhörpositionen sowie für vier Lautsprecher und vier Abhörpositionen durchgeführt.

Robuste Raumhallkompensation durch perturbierte Instanzen der Kanäle		
Parameter	nPRQ in dB	SFM
Ungeformt	9.9	0.63
nicht robust, $\alpha = 40$	10.5	0.64
robust, $\widehat{R} = 4$, $D = 0.5\,\mathrm{cm}$, $\alpha = 10$, $M_{\mathrm{dB}} = 10\,\mathrm{dB}$	7.6	0.75
robust, $\widehat{R} = 4$, $D = 2\,\mathrm{cm}$, $\alpha = 2$, $M_{\mathrm{dB}} = 5\,\mathrm{dB}$	5.8	0.61
robust, $\widehat{R} = 14$, $D = 0.5\,\mathrm{cm}$, $\alpha = 2$, $M_{\mathrm{dB}} = 5\,\mathrm{dB}$	4.5	0.67
robust, $\widehat{R} = 14$, $D = 1\,\mathrm{cm}$, $\alpha = 1$, $M_{\mathrm{dB}} = 25\,\mathrm{dB}$	4.1	0.59
robust, $\widehat{R} = 29$, $D = 1\,\mathrm{cm}$, $\alpha = 2$, $M_{\mathrm{dB}} = 20\,\mathrm{dB}$	3.7	0.69
robust, $\widehat{R} = 29$, $D = 2\,\mathrm{cm}$, $\alpha = 0$, $M_{\mathrm{dB}} = 15\,\mathrm{dB}$	3.3	0.49

Tabelle 6.14: Ergebnisse für den robusten Filterentwurf durch Optimierung von (4.107). α ist der Gewichtsfaktor für den auf der p-Norm der Betragsfrequenzgänge der Gesamtsysteme basierenden Regularisierungsterm, D ist die angenommene Fehlplatzierung, $\widehat{R}$ ist die Anzahl der künstlich erzeugten Instanzen der akustischen Kanäle und M_{dB} bezeichnet den normalisierten Systemfehler (4.138). Die Maße wurden jeweils über die 40 Testpositionen gemittelt.

Die Vorfilter werden für jedes der Q Quellensignal individuell entworfen (siehe Abschnitt 4.1.2.1). Exemplarisch wurden in den Experimenten die Vorfilter für das jeweils für das rechte Ohr bestimmte Quellensignal entworfen. Die Vorfilter wurden jeweils mit einer Länge von $L_h = 5000$ Zeitkoeffizienten entworfen. Als Gewichtsfenster wurden die in Abschnitt 4.4 definierten Gewichtsfenster für die gewünschten Signalpfade und die Signalpfade der Übersprechkomponenten verwendet. Für das Gewichtsfenster der Übersprechkomponente (4.145) wurde $w_0 = 180$ gewählt, was einer gewünschten Unterdrückung des Übersprechens um $20\log_{10}(180) \approx 45.1\,\mathrm{dB}$ entspricht.

Als weitere Parameter wurden $p_d = 20$, $p_u = 10$, $p_f = 8$ sowie $\alpha = 10$ (Optimierung der Betragsfrequenzgänge der Gesamtsysteme) gewählt.

In Tabelle 6.15 sind die Werte für das nPRQ und das DSCR Maß für die verwendeten Konfigurationen an der Referenzposition angegeben. Algorithmus A (Alg. A in Tabelle 6.15) bezeichnet hierbei den auf der p-Norm basierenden Algorithmus zur Formung nach (4.48). Zum Vergleich wurden die Experimente zusätzlich mit einem quadratischen Gütekriterium ($p_d = p_u = 2$) durchgeführt. Bei Optimierung der quadratischen Norm wurde der Betragsfrequenzgang der globalen Impulsantworten mit einem linearen Prädiktorfehlerfilter entzerrt [85].

In Abbildung 6.25 ist der gewünschte Signalpfad $g_s(n)$ sowie der Signalpfad der Übersprechkomponente $g_c(n)$ bei Verwendung von zwei Lautsprechern und einem

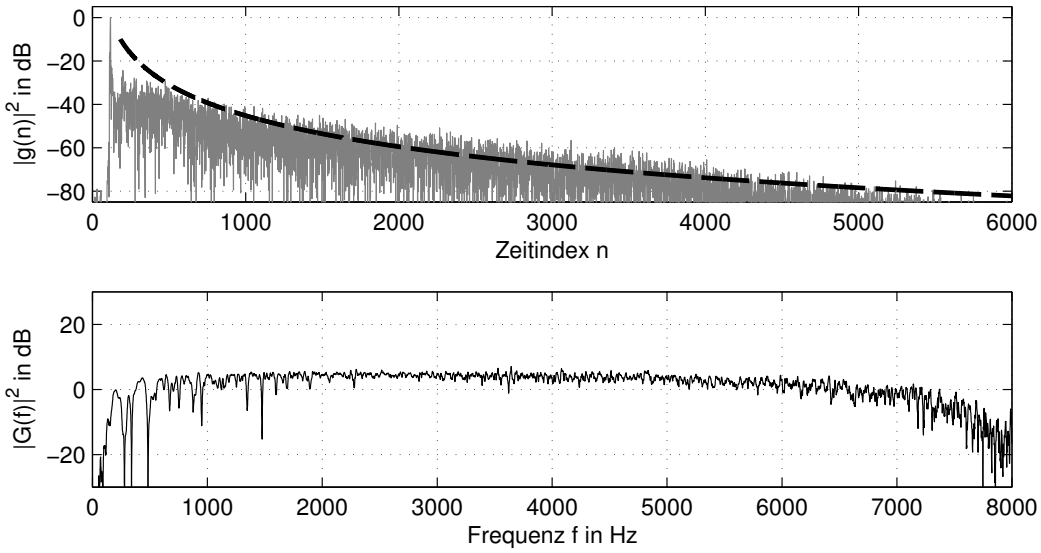

Abbildung 6.24: Globale Impulsantwort an einer der Testpositionen beim robusten Filterentwurf nach (4.107). Zum Filterentwurf wurden $\widehat{R} = 29$ perturbierte Instanzen der von den vier Lautsprechern zur Referenzposition gemessenen Raumimpulsantworten verwendet. Die Parameter für die Generierung der Perturbationen waren $D = 1\,\mathrm{cm}$, $M_{\mathrm{dB}} = 20\,\mathrm{dB}$; der auf der p-Norm basierende Regularisierungsterm zur Optimierung des Betragsfrequenzgangs wurde mit $\alpha = 2$ gewichtet.

Kunstkopf in der Referenzposition dargestellt. Die Vorfilter wurden durch Optimierung der p-Norm (nicht robust) für die Referenzposition entworfen.

In Tabelle 6.16 sind die Ergebnisse zur gemeinsamen Übersprech- und Raumhallkompensation im Falle von räumlicher Fehlplatzierung dargestellt. Hierfür wurde nach dem Filterentwurf das durchschnittliche nPRQ und das durchschnittlicher DSCR Maß über die 40 Testpositionen innerhalb des Abhörvolumens ermittelt. In der Tabelle bezeichnet Algorithmus A den nicht robusten Filterentwurf für die zentrale Referenzposition innerhalb des Abhörvolumens. Algorithmus B bezeichnet den robusten Filterentwurf

Gemeinsame Raumhall- und Übersprechkompensation an der Referenzposition								
Lautsprecher	2		4		4		4	
Mikrofone	2		2		4 (Kopf 1)		4 (Kopf 2)	
Maße in dB	nPRQ	DSCR	nPRQ	DSCR	nPRQ	DSCR	nPRQ	DSCR
Ungeformt	8.4	6.1	8.4	6.1	8.4	6.1	10.3	10.7
Alg. A (p-norm)	0.6	41.6	0.0	51.7	4.2	36.7	0.1	43.7
Alg. A (2-norm)	4.6	18.7	1.6	35.6	5.6	23.2	4.6	25.7

Tabelle 6.15: Ergebnisse für die gemeinsame Raumhall- und Übersprechkompensation bei einem nicht-robusten Filterentwurf an der Referenzposition.

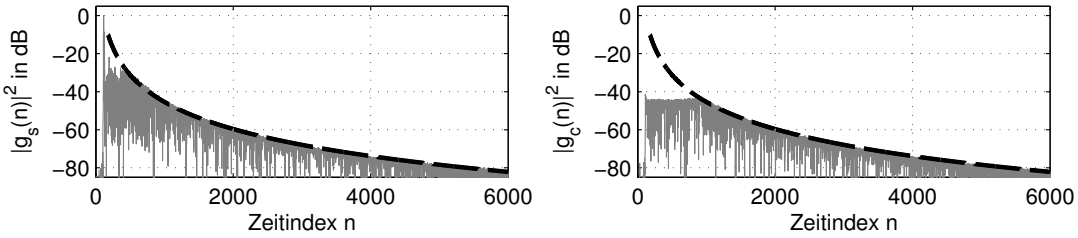

Abbildung 6.25: Gewünschter Signalpfad ($g_s(n)$, links) und Signalpfad der Übersprechkomponente ($g_c(n)$, rechts) an der Referenzposition auf einer logarithmischen Skala. Für den Filterentwurf wurde die p-Norm der Impulsantworten an der Referenzposition optimiert.

Gemeinsame Raumhall- und Übersprechkompensation bei Fehlplatzierung								
Lautsprecher	2		4		4		4	
Mikrofone	2		2		4 (Kopf 1)		4 (Kopf 2)	
Maße in dB	nPRQ	DSCR	nPRQ	DSCR	nPRQ	DSCR	nPRQ	DSCR
Alg. A (p-norm)	8.2	25.5	9.1	22.7	9.6	23.6	7.2	28.9
Alg. B (p-norm)	9.2	26.8	4.6	32.2	10.8	24.2	4.7	32.5
Alg. C (p-norm)	7.2	25.7	7.2	23.1	11.3	19.7	5.7	28.8
Alg. A (2-norm)	11.2	17.7	13.6	18.1	12.6	19.9	10.8	23.8
Alg. B (2-norm)	6.6	20.4	8.4	14.7	8.2	14.1	8.7	18.0
Alg. C (2-norm)	11.8	22.5	13.0	17.3	10.7	19.5	9.2	21.2

Tabelle 6.16: Ergebnisse der gemeinsamen Raumhall- und Übersprechkompensation im Falle von räumlicher Fehlplatzierung. Dargestellt sind Ergebnisse für den nicht-robusten Entwurf (Alg. A), den robusten Entwurf durch Abtastung des Abhörvolumens (Alg. B) und für den robusten Filterentwurf durch statistische Fehlermodellierung (Alg. C). Die Maße wurden jeweils über die 40 Testpositionen gemittelt.

auf Basis der Abtastung des Abhörvolumens. Algorithmus C bezeichnet den räumlich robusten Filterentwurf durch die statistische Fehlermodellierung aus Abschnitt 4.3.1. Als Gewichtsfaktor für den Fehlerterm (4.112) wurde $\beta = 10^{-4}$ verwendet.

Auch in Tabelle 6.16 wird wieder die Formung durch Optimierung der p-Norm für $p_d = 20$ und $p_u = 10$ mit der Formung durch die Optimierung einer quadratischen Norm verglichen. Bei Optimierung der quadratischen Norm wurde der Betragsfrequenzgang der globalen Impulsantworten durch ein lineares Prädiktorfehlerfilter entzerrt [85].

In Abbildung 6.26 sind die Übertragungskanäle des gewünschten Signals sowie der Übersprechkomponente im Falle von räumlicher Fehlplatzierung innerhalb des Abhörvolumens dargestellt; es wurde jeweils die p-Norm der Zeitbereichsdarstellung mit $p_d = 20$ und $p_u = 10$ optimiert. In der oberen Reihe ist das Ergebnis bei einem nicht-robusten Filterentwurf dargestellt. In der mittleren Reihe ist das Ergebnis bei einem robusten

Entwurf durch Abtastung des Abhörvolumens nach [114] und in der unteren Reihe ist das Ergebnis bei einem robusten Entwurf durch die statistische Fehlermodellierung (Abschnitt 4.3.1) dargestellt .

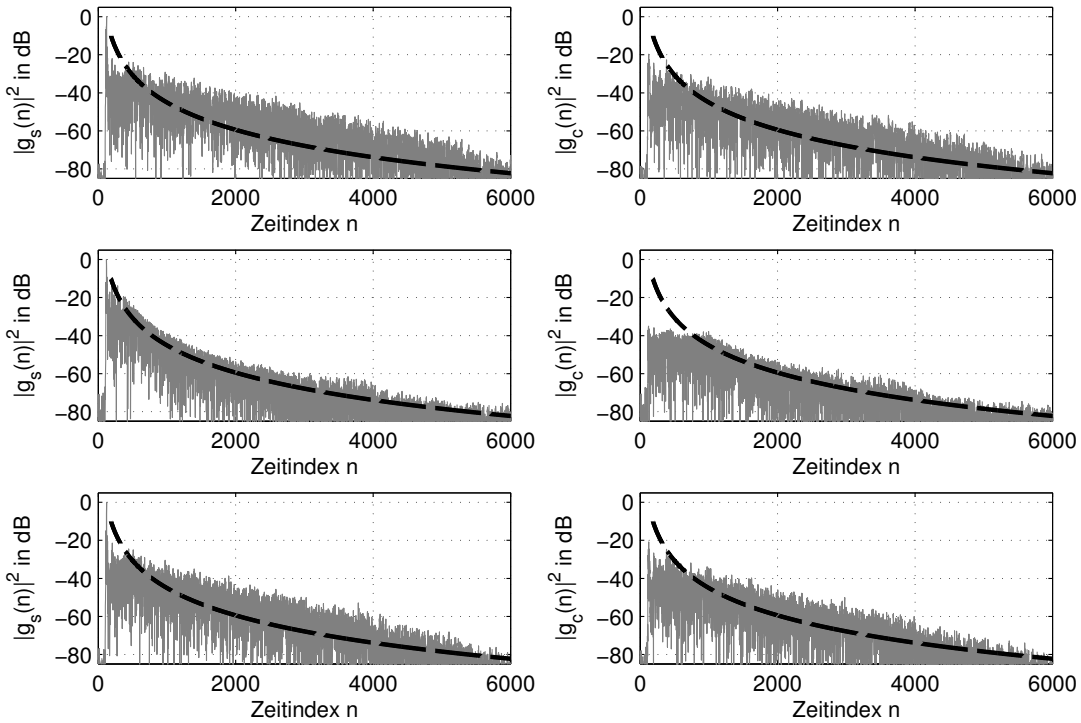

Abbildung 6.26: Gewünschter Signalpfad ($g_s(n)$, links) und Signalpfad der Übersprechkomponente ($g_c(n)$, rechts) auf einer logarithmischen Skala bei Fehlplatzierung innerhalb des Abhörvolumens. Dargestellt ist der nicht-robuste Entwurf (obere Reihe), der robuste Entwurf durch Abtastung des Abhörvolumens (mitte) und der robuste Entwurf durch statistische Fehlermodellierung (unten).

Zusammenfassung - Übersprechkompensation In diesem Abschnitt wurden die Ergebnisse der gemeinsamen Nachhall- und Übersprechkompensation präsentiert. Der Entwurf der Filter erfolgte mit dem in dieser Arbeit entwickelten Rahmenwerk. Die Ergebnisse der Untersuchungen bezüglich der gemeinsamen Kompensation des wahrnehmbaren Nachhalls und der Übersprechkompensation im Falle von räumlicher Fehlplatzierung stehen im Einklang mit den in den Abschnitten 6.2.2.1 (Robustheit durch Abtastung des Abhörvolumens) und 6.2.2.2 (Robustheit durch statistische Modellierung des Fehlers). Die Enthallung und Übersprechkompensation im Falle von räumlicher Fehlplatzierung ist bei dem Filterentwurf nach dem Abtasttheorem für Raumimpuls-

antworten dem robusten Filterentwurf durch statistische Modellierung überlegen (siehe Tabelle 6.16 sowie Abbildung 6.26).

6.3 Zusammenfassung

In diesem Kapitel wurden die mit den neuen Methoden erzielten Ergebnisse präsentiert. Der Schwerpunkt liegt hierbei auf der gleichzeitigen Formung der Zeitbereichsdarstellung und der Frequenzdarstellung des Gesamtsystems sowie auf der räumlich robusten Formung. Des Weiteren wurden Ergebnisse für die gemeinsame Übersprechkompensation und Raumhallkompensation präsentiert. Die Ergebnisse wurden zur besseren Übersicht am Ende jedes Abschnittes kurz bewertet.

Eine Teilmenge der vorgestellten Ergebnisse wurde in [73–81] veröffentlicht.

7 Diskussion und Fazit

In diesem Kapitel werden die in Kapitel 4 erarbeiteten neuen Methoden zur kanalbasierten Raumhallkompensation diskutiert. Hierzu werden die in Kapitel 6 präsentierten Ergebnisse bewertet.

7.1 Diskussion

Die im Rahmen dieser Arbeit entwickelten kanalbasierten Verfahren erlauben den Entwurf von Filtern die sowohl zur Vor- als auch zur Nachfilterung von Audiosignalen verwendet werden können. Im Gegensatz dazu eignen sich nichtlineare Verfahren wie die *spektrale Subtraktion* lediglich zur Nachbearbeitung von aufgenommenen Signalen [53, 92, 99].

Gemeinsame Raumhall- und Übersprechkompensation Im Rahmen dieser Arbeit wurde erstmalig ein Formelwerk zur gemeinsamen Übersprech- und Raumhallkompensation entwickelt [79, 80]. Die erarbeiteten Verfahren basieren auf der Optimierung einer auf der p-Norm basierenden Zielfunktion. Durch eine geeignete Wahl der Gewichtsfenster kann der temporale Maskierungseffekt des menschlichen auditorischen Systems ausgenutzt werden, um eine Reduktion des wahrnehmbaren Nachhalls zu erzielen [35, 120].

Im Gegensatz zur Optimierung einer quadratischen Norm weist das Ergebnis bei der Optimierung einer p-Norm bei einer geeigneten Wahl von p (üblicherweise $p \geq 10$) keinen diffusen späten Nachhall auf [116, 120]. Ersichtlich ist dies anhand der Abbildungen 3.10, 3.12 und 3.13, in denen jeweils die quadratische Norm des ungewünschten Teils minimiert wurde, im Vergleich zu dem Ergebnis der Optimierung mit $p_u = 10$ (Abbildung 3.16). Des Weiteren wird durch die Optimierung der p-Norm eine gleichmäßige Formung der globalen Impulsantwort erreicht. Während in Abbildung 3.18 das Ergebnis der Optimierung einer quadratischen Norm dargestellt ist, wird in Abbildung 3.17 das Ergebnis der Optimierung mit $p_d = 20$, $p_u = 10$ dargestellt. Deutlich ist zu

erkennen, dass das Abklingverhalten der globalen Impulsantwort bei Optimierung mit $p \geq 10$ gleichmäßig dem gewünschten Abklingverhalten folgt.

In Abschnitt 6.2.3 wurde die gemeinsame Enthallung und Übersprechkompensation untersucht. Hierfür wurden Filter durch Optimierung der p-Norm mit $p_d = 20, p_u = 10$ entworfen und mit dem Ergebnis bei Optimierung einer quadratischen Norm ($p_d = p_u = 2$) verglichen. Es wurde gezeigt, dass das quadratische Gütekriterium in nahezu allen untersuchten Konfigurationen bezüglich der Enthallung und Übersprechkompensation (quantifiziert in nPRQ und DSCR [79]) zu schlechteren Ergebnissen als die allgemeine p-Norm führt; die entsprechenden Ergebnisse sind in Tabelle 6.15 für die Positionierung an der Referenzposition sowie in Tabelle 6.16 im Falle von räumlicher Fehlplatzierung dargestellt.

Die im Rahmen dieser Arbeit entwickelte Methode zur gemeinsamen Übersprech- und Raumhallkompensation durch die Impulsformung wurde von Betlehem et al. in [11] aufgegriffen. Betlehem et al. verwendeten eine auf der Maximumsnorm basierende Zielfunktion, die ein gegebenes Zielsystem approximiert. Im Gegensatz zu den hier entwickelten Methoden werden von Betlehem et al. keine robusten Filter entworfen.

In [44, 46] sowie [166] wird die subjektiv wahrgenommene Qualität bei Anwendung unterschiedlicher Verfahren zur Raumhallkompensation untersucht. Es wurde gezeigt, dass die Formung durch die Optimierung einer allgemeinen p-Norm zu einer qualitativ hochwertigen Enthallung führt.

In [42, 44, 46] wird darüber hinaus untersucht, inwiefern die objektiven Maße zur Bewertung eines akustischen Kanals mit der subjektiv wahrgenommenen Qualität korrelieren. In [42] wurde gezeigt, dass das im Verlauf dieser Arbeit entwickelte Maß zur Quantifizierung des wahrnehmbaren Nachhalls [81] eine hohe Korrelation zum subjektiv wahrgenommenen Nachhall sowie zur wahrgenommenen Gesamtqualität besitzt; das im Rahmen dieser Arbeit entwickelte und in [79] publizierte nPRQ Maß ist eine normalisierte Version des in [81] eingeführten Maßes.

Regularisierung des Spektralbereiches Im Rahmen dieser Arbeit wurden zwei Verfahren zur gemeinsamen Optimierung der Zeit- und Frequenzbereichsdarstellung des akustischen Kanals entwickelt. Der Vergleich von Abbildung 6.5 und Abbildung 6.6 zeigt, dass eine gemeinsame Optimierung von Zeit- und Frequenzbereichsdarstellung einer Nachfilterung zur Entzerrung des Frequenzgangs vorzuziehen ist. Schon *kurze* Postfilter zur Entzerrung des Frequenzgangs wirken sich negativ auf die Formung im Zeitbereich aus.

Ein Vergleich der Ergebnisse für die mit dem Kunstkopf gemessene Raumimpulsantwort (Tabellen 6.1, 6.2, 6.7), für die im TG Oldenburg gemessene Raumimpulsantwort (Tabellen 6.3, 6.4, 6.8) und für die im Wohn- und Esszimmer gemessene Raumimpulsantwort (Tabellen 6.5, 6.6, 6.9) zeigt, dass die Regularisierung durch das gSFM im Allgemeinen zu besseren Ergebnissen bezüglich des nPRQ und SFM Maßes führt als die Regularisierung durch die p-Norm des Betragsfrequenzgangs.

Des Weiteren erlaubt die Regularisierung auf Grundlage des generalisierten Maßes zur Bewertung der spektralen Flachheit die Eigenschaften der zu übertragenden Signale beziehungsweise der zur Reproduktion verwendeten Komponenten zu berücksichtigen. In Abbildung 6.11 ist das Ergebnis eines auf die zur Reproduktion verwendeten Komponenten angepassten Filterentwurfs zu sehen. Ein Vergleich von Tabelle 6.9 mit Tabelle 6.10 zeigt, dass eine auf die Wiedergabemodalitäten angepasste Regularisierung zu einer verbesserten Enthallung führt.

Das im Rahmen dieser Arbeit erarbeitete generalisierte Maß zur Bewertung der spektralen Flachheit (siehe Abschnitt 4.2.2.1) ermöglicht es, beliebige Anforderungen an den Betragsfrequenzgang des Gesamtsystems zu formulieren. Es wurde gezeigt, wie ein *flacher* Betragsfrequenzgang gefordert werden kann. Des Weiteren wurde gezeigt, wie der durch ein Audiogramm gemessene frequenzabhängige Hörverlust während der Impulsformung kompensiert werden kann. Auch für diesen Fall ist der Vergleich zu der nachträglichen Entzerrung des Betragsfrequenzgangs gegeben. Der Vergleich der Abbildungen 6.15 und 6.13 zeigt, dass auch für dieses Szenario die gleichzeitige Optimierung einer Nachfilterung vorzuziehen ist.

In einem weiteren Experiment wurde gezeigt, wie die Integration in Teilbänder zur Optimierung des Betragsfrequenzgangs genutzt werden kann (Abbildung 6.17). Für das Experiment wurde der Betragsfrequenzgang in 27 gehörrichtige Teilbänder auf Grundlage einer Gammatone-Filterbank zusammengefasst. Durch die Integration werden die Anforderungen an die Frequenzbereichsdarstellung *abgeschwächt*: Es wird nur noch die Gesamtenergie eines Bandes und nicht jeder einzelne Koeffizient bewertet. Dementsprechend kann eine verbesserte Enthallung erreicht werden. Ein Vergleich von Abbildung 6.15 und 6.17 verdeutlicht diesen Zusammenhang.

Räumliche Robustheit Im Rahmen dieser Arbeit wurden zwei neue Verfahren zum Entwurf räumlich robuster Filter zur Impulsformung entwickelt. Während die erste Methode (siehe Abschnitt 4.3.1) auf einem statistischen Fehlermodell basiert werden für die zweite vorgeschlagene Methode synthetische Raumimpulsantworten in der direkten

Umgebung zur Referenzposition generiert; hierfür wurde ein entsprechendes Verfahren erarbeitet (Abschnitt 4.3.2).

Für beide Verfahren wurde experimentell nachgewiesen, dass der räumliche Gültigkeitsbereich in dem eine Raumhallkompensation gewährleistet ist, vergrößert werden konnte.

Während der nicht robuste Entwurf schon bei kleinsten Bewegungen zu einer Erhöhung des wahrnehmbaren Nachhalls führt, ermöglichen die erarbeiteten Verfahren Bewegungen innerhalb des Abhörvolumens bei reduziertem Nachhall. Exemplarisch ist in Abbildung 6.19 die globale Impulsantwort abseits der Referenzposition bei einem nicht-robusten Filterentwurf dargestellt. In Tabelle 6.11 ist ersichtlich, dass der wahrnehmbare Nachhall über die 40 Testpositionen gemittelt bei einem nicht-robusten Filterentwurf von $\mathrm{nPRQ} = 9.9\,\mathrm{dB}$ auf $\mathrm{nPRQ} = 10.5\,\mathrm{dB}$ ansteigt.

Das auf dem statistischen Fehlermodell basierende Verfahren ist sehr speicherintensiv, da für jeden akustischen Kanal eine Matrix der Größe $L_h \times L_h$ im Speicher vorgehalten werden muss; L_h notiert hierbei die Länge der zu entwerfenden Filter. Der zusätzliche Speicheraufwand für das auf den perturbierten Instanzen basierende Verfahren ist zu vernachlässigen ($\widehat{R} \cdot L_c$, wobei $\widehat{R}$ die Anzahl an Perturbationen pro Kanal ist und L_c die Länge der Raumimpulsantwort). Des Weiteren zeigt ein Vergleich der Werte in den Tabellen 6.12 und 6.14, dass der robuste Filterentwurf auf Basis der perturbierten Raumimpulsantworten zu vergleichbaren Ergebnissen bezüglich der Reduktion des wahrnehmbaren Nachhalls führt.

7.2 Zusammenfassung und Fazit

Die im Rahmen dieser Arbeit neu entwickelten Verfahren sind essentiell für die praktische Nutzbarmachung kanalbasierter Verfahren zur Raumhallkompensation. Durch die Vorfilterung der Audiosignale zur Raumhall- und Übersprechkompensation vor der Wiedergabe über Lautsprecher kann ein qualitativ hochwertiges Hörerlebnis ermöglicht werden. Die entworfenen Filter können darüber hinaus auch zur Nachbearbeitung auf der Empfängerseite verwendet werden, um beispielsweise die Erkennungsraten eines Systems zur automatischen Spracherkennung zu erhöhen.

Im Rahmen dieser Arbeit wurde ein verallgemeinertes Formelwerk zur mehrkanaligen Impulsformung durch die Optimierung einer auf der p-Norm basierenden Zielfunktion entwickelt. Es wurde gezeigt, wie sich unterschiedliche Anwendungsszenarien in dem Formelwerk darstellen und optimieren lassen. Des Weiteren wurde gezeigt, wie das

Problem der Übersprechkompensation in dem Formelwerk ausgedrückt und gelöst werden kann.

Es wurde gezeigt, dass die gleichzeitige Optimierung der Zeit- und Frequenzbereichsdarstellung des Gesamtsystems grundsätzlich einer Verarbeitung in zwei Schritten (Schritt 1. Optimierung der Zeitbereichsdarstellung, Schritt 2. Approximation eines gewünschten Betragsfrequenzgangs) vorzuziehen ist. Hierfür wurden zwei Optimalitätskriterien vorgestellt. Während das erste Verfahren die p-Norm des Betragsfrequenzgangs optimiert, wird für das zweite Verfahren eine Optimierung des neu eingeführten generalisierten Maßes zur Bewertung der spektralen Flachheit (gSFM) durchgeführt. Es konnte darüber hinaus gezeigt werden, dass die Optimierung des gSFM Maßes bei geeigneter Parametrierung zu einem Gesamtsystem mit beliebigem vorgegebenen Betragsfrequenzgang führt.

Des Weiteren wurden zwei neue Methoden vorgestellt, um den räumlichen Gültigkeitsbereich der Formung zu vergrößern. Der Mehrpositions-Ansatz, welcher auf der Messung vieler Raumimpulsantworten innerhalb des Abhörvolumens basiert, ist oftmals nicht praktikabel, da die Messungen von Raumimpulsantworten auf einem feinen räumlichen Gitter sehr zeitaufwendig sind. Die erarbeiteten Verfahren ermöglichen eine räumlich robuste Impulsformung, wobei nur die akustischen Kanäle zu einer Referenzposition gemessen werden müssen.

Das erste erarbeitete Verfahren basiert auf der Minimierung eines quadratischen Fehlerterms, der die statistischen Eigenschaften der Systemfehler im Falle von räumlicher Fehlplatzierung abbildet. Hierfür muss für jede gemessene Raumimpulsantwort eine Matrix aufgestellt werden. Es wurde die Effektivität dieses Verfahrens demonstriert und es wurde ein Algorithmus entwickelt, um die Berechnung der benötigten Matrizen um mehrere Größenordnungen zu beschleunigen. Das zweite erarbeitete Verfahren basiert auf künstlich perturbierten Instanzen der gemessenen Raumimpulsantworten. Anhand eines neu entwickelten Modells werden perturbierte Instanzen der akustischen Kanäle generiert und gleichzeitig mit der Referenzmessung zusammen geformt. Der perturbationsbasierte Ansatz ist generell dem auf dem statistischen Fehlermodell basierenden Ansatz überlegen: Es müssen keine *großen* Matrizen im Speicher gehalten werden, die Generierung der perturbierten Instanzen ist *schnell* und die Ergebnisse sind bezüglich der zur Bewertung verwendeten Maße (nPRQ und SFM) gleichwertig.

Während beim nicht-robusten Filterentwurf lediglich der akustische Kanal an der Referenzposition geformt wird, erlauben die neuen Verfahren die Bewegung des Zuhörers

innerhalb eines kleinen Raumvolumens, ohne dass es zu einer Beeinträchtigung der wahrgenommenen Qualität kommt.

8 Ausblick

Die im Rahmen dieser Arbeit entwickelten Verfahren stellen einen wichtigen Schritt zur Nutzbarmachung von kanalbasierten Verfahren zur Raumhallkompensation und Übersprechkompensation in akustischen Szenarien dar. Dessen ungeachtet ist es notwendig, sowohl die Robustheit gegenüber Fehlplatzierungen als auch die Geschwindigkeit des Filterentwurfs weiter zu erhöhen.

Ein möglicher Weg für einen schnelleren Filterentwurf wurde von Mazur et al. in [110] aufgezeigt. Hier wurde der Algorithmus von Mei et al. [114] zum robusten Filterentwurf durch Abtastung eines Volumens auf einer Grafikkarte [87] implementiert. Hierzu wurde die CUDA (englisch: *Compute Unified Device Architecture*) Schnittstelle [143] verwendet.

Auf algorithmischer Ebene wurde von Mazur et al. in [111] vorgeschlagen, die Lernregel des Gradientenabstiegsverfahrens zur Optimierung der auf der p-Norm basierenden Zielfunktion aus [120] zu modifizieren. Hierzu wird der Gradient bezüglich der Filterkoeffizienten mit einer Fensterfunktion skaliert, die spätere Koeffizienten weniger stark gewichtet als frühe. Die Motivation hierfür ist, den Gradienten derart zu modifizieren, dass er das *fertige* Filter in Bezug auf das Abklingverhalten möglichst gut approximiert. Ein ähnliches Verfahren wurde beispielsweise in [48, 107] zur schnelleren Identifikation eines akustischen Kanals verwendet; dort wird der Ansatz auch als *Proportionate-LMS* bezeichnet [29].

Die im Rahmen dieser Arbeit entwickelten Verfahren arbeiten nicht blind, sondern basieren auf einer Schätzung der akustischen Kanäle. In Verbindung mit einem System zur blinden Kanalidentifikation (siehe Abschnitt 2.1.3.2 beziehungsweise [54, 56, 58, 65–67]) können die im Rahmen dieser Arbeit entwickelten Verfahren allerdings zu *blinden* Verfahren weiterentwickelt werden.

Eine weitere Verbesserung der räumlichen Robustheit ist durch die Verfeinerung der statischen Modelle für die Störterme zu erwarten. Hierbei könnte beispielsweise Vorwissen über die Raumgeometrie und über die tatsächliche Bewegung des Zuhörers (beispielsweise in Echtzeit durch Headtracking ermittelt) in die Modelle integriert werden.

Für die praktische Nutzbarmachung der Mehrpositions-Ansätze ist eine Reduzierung des Messaufwandes nötig. Erste Arbeiten hierzu basieren auf der komprimierten Abtastung der plenakustischen Funktion [1, 122–124]. Hierbei werden die akustischen Kanäle an beliebigen Positionen innerhalb eines großen Volumens auf Grundlage weniger Messungen geschätzt.

Die Enthallung selber könnte durch die Integration von neuen psychoakustischen Erkenntnissen weiter verbessert werden. So wird im Rahmen dieser Arbeit für die komplette Bandbreite die durchschnittliche temporale Maskierungsschwelle nach Fielder et al. [35] angenommen. Es ist zu erwarten, dass unterschiedliche Bänder unterschiedliche Maskierungskurven aufweisen. Die hier entwickelten Verfahren könnten auf einzelne Teilbänder angewendet werden. In [168] wurde durch die Dekomposition der Raumimpulsantworten in unterabgetastete Teilbänder eine Reduzierung des Berechnungsaufwandes zum Filterentwurf erreicht.

In [42] wurde gezeigt, dass die Impulsformung mit einem dem nPRQ Maß entsprechenden Gewichtsfenster zu einer verbesserten subjektiv wahrgenommenen Qualität führt.

Abschließend muss festgestellt werden, dass für die Weiterentwicklung der erarbeiteten Verfahren formelle Hörtests mit einer ausreichend großen Testgruppe essentiell sind. Hörtests sind des Weiteren nötig, um die Gewichtsfenster bezüglich der wahrgenommenen Gesamtqualität zu optimieren.

Literatur

[1] Thibaut Ajdler, Luciano Sbaiz und Martin Vetterli. „The Plenacoustic Function and Its Sampling“. In: *IEEE Transactions on Signal Processing* 54.10 (Okt. 2006), S. 3790–3804. DOI: 10.1109/TSP.2006.879280.

[2] Jont B. Allen und David A. Berkley. „Image Method for Efficiently Simulating Small-Room Acoustics“. In: *Journal of the Acoustical Society of America* 65.4 (Apr. 1979), S. 943–950. DOI: 10.1121/1.382599.

[3] Herbert Amann und Joachim Escher. *Analysis III*. 1. Aufl. Birkhäuser Verlag, 2001.

[4] Güner Arslan, Brian L. Evans und Sayfe Kiaei. „Optimum Channel Shortening for Discrete Multitone Transceivers“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Bd. 5. Istanbul, Turkey, Juni 2000, S. 2965–2968. DOI: 10.1109/ICASSP.2000.861156.

[5] Niklas Bährecke. „Acoustic Crosstalk Cancellation with Room Modelling“. Bachelorarbeit, Institut für Signalverarbeitung, Universität zu Lübeck. Nov. 2012.

[6] Jacob Benesty u. a. *Advances in Network and Acoustic Echo Cancellation*. Berlin: Springer, 2001.

[7] Jacob Benesty u. a. „Stereophonic Acoustic Echo Cancellation Using Nonlinear Transformations and Comb Filtering“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Seattle, Washington, USA, Mai 1998, S. 3673–3676. DOI: 10.1109/ICASSP.1998.679680.

[8] Augustinus J. Berkhout. „A Holographic Approach to Acoustic Control“. In: *Journal of the Audio Engineering Society* 36.12 (Dez. 1988), S. 979–995.

[9] Augustinus J. Berkhout, Diemer de Vries und Marinus M. Boone. „A New Method to Acquire Impulse Responses in Concert Halls“. In: *Journal of the Acoustical Society of America* 68.1 (Juli 1980), S. 179–183. DOI: 10.1121/1.384618.

[10] Augustinus J. Berkhout, Diemer de Vries und Peter Vogel. „Acoustic Control by Wave Field Synthesis“. In: *Journal of the Acoustical Society of America* 93.5 (Mai 1993), S. 2764–2778. DOI: 10.1121/1.405852.

[11] Terence Betlehem, Paul D. Teal und Yusuke Hioka. „Efficient Crosstalk Canceler Design with Impulse Response Shortening Filters“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Kyoto, Japan, März 2012, S. 393–396. DOI: 10.1109/ICASSP.2012.6287899.

[12] Sunil Bharitkar, Philip Hilmes und Chris Kyriakakis. „Robustness of Spatial Average Equalization: A Statistical Reverberation Model Approach“. In: *Journal of the Acoustical Society of America* 116.6 (Dez. 2004), S. 3491–3497. DOI: 10.1121/1.1810238.

[13] Jens Blauert. *The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press, 1983.

[14] Stephen Boyd und Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[15] Aditya Chopra und Brian L. Evans. „Design of Sparse Filters for Channel Shortening“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Dallas, Texas, USA, März 2010, S. 1518–1521. DOI: 10.1109/ICASSP.2010.5495507.

[16] W. T. Chu. „Impulse-Response and Reverberation-Decay Measurements Made by Using a Periodic Pseudorandom Sequence“. In: *Applied Acoustics* 29 (1990), S. 193–205.

[17] Andrzej Cichocki und Shun-Ichi Amari. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, LTD, 2002.

[18] Duane H. Cooper und Jerald L. Bauck. „Prospects for Transaural Recording“. In: *Journal of the Audio Engineering Society* 37.1/2 (Jan. 1989), S. 30–33.

[19] Henry Cox. „Linear versus Logarithmic Averaging“. In: *Journal of the Acoustical Society of America* 39.4 (1966), S. 688–690. DOI: 10.1121/1.1909942.

[20] Donnacha Daly, Conor Heneghan und Anthony D. Fagan. „A Minimum Mean-Squared Error Interpretation of Residual ISI Channel Shortening for Discrete Multitone Transceivers“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Bd. 4. Salt Lake City, Utah, USA, Mai 2001, S. 2065–2068. DOI: 10.1109/ICASSP.2001.940398.

[21] P. Damaske. „Head-Related Two-Channel Stereophony with Loudspeaker Reproduction“. In: *Journal of the Acoustical Society of America* 50 (1971), S. 1109–1115. DOI: 10.1121/1.1912742.

[22] Jérôme Daniel. „Acoustic Field Representation, Application to the Transmission and the Reproduction of Complex Sound Environments in a Multimedia Context“. Diss. Université Paris, 2000.

[23] Jérôme Daniel. „Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format“. In: *23rd International Conference on Multichannel Audio, Audio Engineering Society*. Copenhagen, Denmark, Mai 2003.

[24] Torsten Dau, Birger Kollmeier und Armin Kohlrausch. „Modelling Auditory Processing of Amplitude Modulation. I. Detection and Masking with Narrow-Band Carriers“. In: *Journal of the Acoustical Society of America* 102.5 (Nov. 1997), S. 2892–2905. DOI: 10.1121/1.420344.

[25] Guillaume Defrance und Jean-Dominique Polack. „Measuring the Mixing Time in Auditoria“. In: *Proc. Acoustics 08 Paris*. Paris, France, Juni 2008, S. 3871–3876. DOI: 10.1121/1.2934368.

[26] Marc Delcroix, Takafumi Hikichi und Masato Miyoshi. „Improved Blind Dereverberation Performance by Using Spatial Information“. In: *Proc. Interspeech 2005*. Lisbon, Portugal, Sep. 2005, S. 2309–2312.

[27] Marc Delcroix, Tomohiro Nakatani und Shinji Watanabe. „Static and Dynamic Variance Compensation for Recognition of Reverberant Speech With Dereverberation Preprocessing“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 17.2 (Feb. 2009), S. 324–334. DOI: 10.1109/TASL.2008.2010214.

[28] Pascal Dietrich u. a. „Time Efficient Measurement Method for Individual HRTFs“. In: *Proceedings of the German Annual Conference on Acoustics (DAGA)*. Darmstadt, Germany, März 2012, S. 333–334.

[29] Donald L. Duttweiler. „Proportionate Normalized Least-Mean-Squares Adaptation in Echo Cancelers“. In: *IEEE Transactions on Speech and Audio Processing* 8.5 (Sep. 2000), S. 508–518. DOI: 10.1109/89.861368.

[30] Stephen J. Elliott und Philip A. Nelson. „Multiple-Point Equalization in a Room Using Adaptive Digital Filters“. In: *Journal of the Audio Engineering Society* 37.11 (Nov. 1989), S. 899–907.

[31] D. D. Falconer und F. R. Magee. „Adaptive Channel Memory Truncation for Maximum Likelihood Sequence Estimation“. In: *The Bell System Technical Journal* 52.9 (Nov. 1973), S. 1541–1562.

[32] Angelo Farina. „Advancements in Impulse Response Measurements by Sine Sweeps“. In: *Proc. 122nd Convention of the Audio Engineering Society*. Vienna, Austria, Mai 2007.

[33] Angelo Farina. „Impulse Response Measurements“. In: *Proceedings of the 23rd Nordic Sound Symposium*. Bolkesjø, Norway, Sep. 2007.

[34] Louis D. Fielder. „Analysis of Traditional and Reverberation-Reducing Methods of Room Equalization“. In: *Journal of the Audio Engineering Society* 51.1 (Jan. 2003), S. 3–26.

[35] Louis D. Fielder. „Practical Limits for Room Equalization“. In: *Proc. 111th Convention of the Audio Engineering Society*. Nov. 2001, S. 1–19.

[36] Ken'ichi Furuya und Akitoshi Kataoka. „Robust Speech Dereverberation Using Multichannel Blind Deconvolution with Spectral Subtraction“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 15.5 (Juli 2007), S. 1579–1591. DOI: 10.1109/TASL.2007.898456.

[37] William G. Gardner. „3-D Audio Using Loudspeakers“. Diss. School of Architecture und Planning, Massachusetts Institute of Technology, Sep. 1997.

[38] William G. Gardner. „Head Tracked 3-D Audio Using Loudspeakers“. In: *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. New Paltz, New York, USA, Okt. 1997.

[39] Timo Gerkmann und Richard C. Hendriks. „Unbiased MMSE-Based Noise Power Estimation with Low Complexity and Low Tracking Delay“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 20.4 (Mai 2012), S. 1383–1393. DOI: 10.1109/TASL.2011.2180896.

[40] Bradford W. Gillespie und Les E. Atlas. „Acoustic Diversity for Improved Speech Recognition in Reverberant Environments“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Orlando, Florida, USA, Mai 2002, S. 557–560. DOI: 10.1109/ICASSP.2002.5743778.

[41] Bradford W. Gillespie und Les E. Atlas. „Strategies for Improving Audible Quality and Speech Recognition Accuracy of Reverberant Speech“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Hong Kong, China, Apr. 2003, S. 676–679. DOI: 10.1109/ICASSP.2003.1198871.

[42] Stefan Goetze u. a. „A Study on Speech Quality and Speech Intelligibility Measures for Quality Assessment of Single-Channel Dereverberation Algorithms“. In: *Proceedings of the IEEE International Workshop on Acoustic Signal Enhancement*. Antibes, France, 2014.

[43] Stefan Goetze u. a. „Estimation of the Optimum Delay for Speech Dereverberation by Inverse Filtering“. In: *Proceedings of the German Annual Conference on Acoustics (DAGA)*. Rotterdam, The Netherlands, März 2009, S. 976–979.

[44] Stefan Goetze u. a. „Quality Assessment for Listening-Room Compensation Algorithms“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Dallas, TX, USA, März 2010, S. 2450–2453. DOI: 10.1109/ICASSP.2010.5496301.

[45] Stefan Goetze u. a. „Spatial Sensitivity for Listening Room Compensation“. In: *Demonstration given at IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2007)*. New Paltz, USA, Okt. 2007.

[46] Stefan Goetze u. a. „Speech Quality Assessment for Listening-Room Compensation“. In: *Proc. AES 38th International Conference*. Piteå, Sweden, Juni 2010.

[47] Stefan Goetze u. a. „System Identification for Multi-Channel Listening-Room Compensation using an Acoustic Echo Canceller“. In: *Proc. Workshop on Hands-Free Speech Communication and Microphone Arrays (HSCMA 2008)*. Trento, Italy, Mai 2008, S. 224–227.

[48] Stefan Goetze u. a. „System Identification of Equalized Room Impulse Responses by an Acoustic Echo Canceller using Proportionate LMS Algorithms“. In: *Proc. 130st Convention of the Audio Engineering Society*. London, UK, Mai 2011.

[49] Gene H. Golub und Charles F. van Loan. *Matrix Computations*. 3. Aufl. Johns Hopkins University Press, 1996.

[50] Thomas Görne. *Tontechnik*. Fachbuchverlag Leipzig, 2006.

[51] Augustine H. Gray und John D. Markel. „A Spectral-Flatness Measure for Studying the Autocorrelation Method of Linear Prediction of Speech Analysis“. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 22.3 (Juni 1974), S. 207–217. DOI: 10.1109/TASSP.1974.1162572.

[52] Emanuël A. P. Habets. „Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement“. Diss. Technische Universiteit Eindhoven, 2007.

[53] Emanuël A. P. Habets. „Single-Channel Speech Dereverberation based on Spectral Subtraction“. In: *Proc. of the 15th Annual Workshop on Circuits, Systems and Signal Processing (ProRISC 2004)*. Veldhoven, Netherlands, Nov. 2004, S. 250–254.

[54] Emanuël A. P. Habets, Jacob Benesty und Patrick A. Naylor. „A Cross-Relation Based Affine Projection Algorithm for Blind SIMO System Identification“. In: *Proceedings of the European Signal Processing Conference*. Barcelona, Spain, Aug. 2011, S. 1743–1747.

[55] Hüseyin Hacıhabiboğlu und Zoran Cvetković. „Multichannel Dereverberation Theorems and Robustness Issues“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 20.2 (Feb. 2012), S. 676–689. DOI: 10.1109/TASL.2011.2164528.

[56] Mohammad Ariful Haque und Md. Kamrul Hasan. „Noise Robust Multichannel Frequency-Domain LMS Algorithms for Blind Channel Identification“. In: *IEEE Signal Processing Letters* 15 (2008), S. 305–308. DOI: 10.1109/LSP.2008.917803.

[57] Mohammad Ariful Haque, Toufiqul Islam und Md. Kamrul Hasan. „Robust Speech Dereverberation Based on Blind Adaptive Estimation of Acoustic Channels“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 19.4 (Mai 2011), S. 775–787. DOI: 10.1109/TASL.2010.2064306.

[58] Md. Kamrul Hasan und Patrick A. Naylor. „Analyzing Effect of Noise on LMS-Type Approaches to Blind Estimation of SIMO Channels: Robustness Issue“. In: *Proceedings of the European Signal Processing Conference*. Florence, Italy, Sep. 2006.

[59] Robert L. Hershey. „Analysis of Difference Between Log Mean and Mean Log Averaging“. In: *Journal of the Acoustical Society of America* 51.4 (1972), S. 1194–1197. DOI: 10.1121/1.1912957.

[60] Takayuki Hidaka, Yoshinari Yamada und Takehiko Nakagawa. „A New Definition of Boundary Between Early Reflections and Late Reverberation in Room Impulse Responses“. In: *Journal of the Acoustical Society of America* 122.1 (Juli 2007), S. 326–332. DOI: 10.1121/1.2743161.

[61] Takafumi Hikichi, Marc Delcroix und Masato Miyoshi. „Inverse Filtering for Speech Dereverberation Less Sensitive to Noise and Room Transfer Function Fluctuations“. In: *EURASIP Journal on Advances in Signal Processing* 2007 (Apr. 2007), 12 pages. DOI: 10.1155/2007/34013.

[62] Takafumi Hikichi, Marc Delcroix und Masato Miyoshi. „Inverse Filtering for Speech Dereverberation Less Sensitive to Noise“. In: *Proceedings of the IEEE International Workshop on Acoustic Echo and Noise Control*. Paris, France, Sep. 2006.

[63] Takafumi Hikichi, Marc Delcroix und Masato Miyoshi. „On Robust Inverse Filter Design for Room Transfer Function Fluctuations“. In: *Proceedings of the European Signal Processing Conference*. Florence, Italy, Sep. 2006.

[64] Volker Hohmann. „Frequency Analysis and Synthesis Using a Gammatone Filterbank“. In: *Acta Acustica united with Acustica* 88 (2002), S. 433–442.

[65] Yiteng Arden Huang und Jacob Benesty. „A Class of Frequency-Domain Adaptive Approaches to Blind Multichannel Identification“. In: *IEEE Transactions on Signal Processing* 51.1 (Jan. 2003), S. 11–24. DOI: 10.1109/TSP.2002.806559.

[66] Yiteng Arden Huang und Jacob Benesty. „Adaptive Blind Channel Identification: Multi-Channel Least Mean Square and Newton Algorithms“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Orlando, Florida, USA, Mai 2002, S. 1637–1640. DOI: 10.1109/ICASSP.2002.5744932.

[67] Yiteng Arden Huang, Jacob Benesty und Jingdong Chen. „Identification of Acoustic MIMO Systems: Challenges and Opportunities“. In: *Signal Processing* 86 (Juni 2006), S. 1278–1295.

[68] Walt Jesteadt, Sid P. Bacon und James R. Lehman. „Forward Masking as a Function of Frequency, Masker Level, and Signal Delay“. In: *Journal of the Acoustical Society of America* 71.4 (Apr. 1982), S. 950–962. DOI: 10.1121/1.387576.

[69] Marco Jeub, Magnus Schäfer und Peter Vary. „A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms“. In: *Proc. of International Conference on Digital Signal Processing (DSP 2009)*. Santorini, Greece, Juli 2009.

[70] James D. Johnston. „Transform Coding of Audio Signals Using Perceptual Noise Criteria“. In: *IEEE Journal on Selected Areas in Communications* 6.2 (Feb. 1988), S. 314–323. DOI: 10.1109/49.608.

[71] Jean-Marc Jot, Laurent Cerveau und Olivier Warusfel. „Analysis and Synthesis of Room Reverberation Based on a Statistical Time-Frequency Model“. In: *Proc. 103rd Convention of the Audio Engineering Society*. Sep. 1997.

[72] Jan Ole Jungmann, Stefan Goetze und Alfred Mertins. „Room Impulse Response Reshaping by p-Norm Optimization Based on Estimates of Room Impulse Responses“. In: *Proceedings of the German Annual Conference on Acoustics (DAGA)*. Düsseldorf, Germany, März 2011, S. 611–612.

[73] Jan Ole Jungmann, Radoslaw Mazur und Alfred Mertins. „A Stochastic Approach for Robust Listening Room Compensation“. In: *Proceedings of the German Annual Conference on Acoustics (DAGA)*. Merano, Italy, März 2013, S. 185–186.

[74] Jan Ole Jungmann, Radoslaw Mazur und Alfred Mertins. „Joint Time-Domain Reshaping and Frequency-Domain Equalization of Room Impulse Responses“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Florence, Italy, Mai 2014, S. 6692–6696. DOI: 10.1109/ICASSP.2014.6854885.

[75] Jan Ole Jungmann, Radoslaw Mazur und Alfred Mertins. „On the Equalization and Reshaping of Room Impulse Responses“. In: *Proceedings of the German Annual Conference on Acoustics (DAGA)*. Oldenburg, Germany, März 2014, S. 497–498.

[76] Jan Ole Jungmann, Radoslaw Mazur und Alfred Mertins. „Optimized Preprocessing for Spatially Robust Room Impulse Response Reshaping“. In: *Proceedings of the IEEE International Workshop on Acoustic Signal Enhancement*. Aachen, Germany, Sep. 2012.

[77] Jan Ole Jungmann, Radoslaw Mazur und Alfred Mertins. „Perturbation of Room Impulse Responses and its Application in Robust Listening Room Compensation“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vancouver, BC, Canada, Mai 2013, S. 433–437. DOI: 10.1109/ICASSP.2013.6637684.

[78] Jan Ole Jungmann, Radoslaw Mazur und Alfred Mertins. „Robust Listening Room Compensation by Optimizing Multiple p-Norm Based Criteria“. In: *Proceedings of the German Annual Conference on Acoustics (DAGA)*. Darmstadt, Germany, März 2012, S. 105–106.

[79] Jan Ole Jungmann u. a. „Combined Acoustic MIMO Channel Crosstalk Cancellation and Room Impulse Response Reshaping“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 20.6 (Aug. 2012), S. 1829–1842. DOI: 10.1109/TASL.2012.2190929.

[80] Jan Ole Jungmann u. a. „Robust Combined Crosstalk Cancellation and Listening-Room Compensation“. In: *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. New Paltz, New York, USA, Okt. 2011, S. 9–12. DOI: 10.1109/ASPAA.2011.6082309.

[81] Jan Ole Jungmann u. a. „Room Impulse Response Reshaping by Joint Optimization of Multiple p-Norm Based Criteria“. In: *Proceedings of the European Signal Processing Conference*. Barcelona, Spain, Aug. 2011, S. 1658–1662.

[82] Markus Kallinger und Alfred Mertins. „A Spatially Robust Least Squares Crosstalk Canceller“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Bd. 1. Apr. 2007, S. 177–180. DOI: 10.1109/ICASSP.2007.366645.

[83] Markus Kallinger und Alfred Mertins. „Impulse Response Shortening for Acoustic Listening Room Compensation“. In: *Proceedings of the IEEE International Workshop on Acoustic Echo and Noise Control*. Eindhoven, The Netherlands, Sep. 2005, S. 197–200.

[84] Markus Kallinger und Alfred Mertins. „Multi-Channel Room Impulse Response Shaping - A Study“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Toulouse, France, Mai 2006, S. 101–104. DOI: 10.1109/ICASSP.2006.1661222.

[85] Markus Kallinger und Alfred Mertins. „Room Impulse Response Shortening by Channel Shortening Concepts“. In: *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove, CA , USA, Okt. 2005, S. 898–902. DOI: 10.1109/ACSSC.2005.1599885.

[86] Keisuke Kinoshita u. a. „Suppression of Late Reverberation Effect on Speech Signal Using Long-Term Multiple-Step Linear Prediction“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 17.4 (Mai 2009), S. 534–545. DOI: 10.1109/TASL.2008.2009015.

[87] David B. Kirk und Wen-Mei W. Hwu. *Programming Massively Parallel Processors: A Hands-on Approach*. 1. Aufl. Morgan Kaufmann, 2010.

[88] Ole Kirkeby und Philip A. Nelson. „Digital Filter Design for Inversion Problems in Sound Reproduction“. In: *Journal of the Audio Engineering Society* 47 (Juli 1999), S. 583–595.

[89] Ole Kirkeby u. a. „Fast Deconvolution of Multichannel Systems using Regularization“. In: *IEEE Transactions on Speech and Audio Processing* 6.2 (März 1998), S. 189–194. DOI: 10.1109/89.661479.

[90] Ina Kodrasi und Simon Doclo. „Robust Partial Multichannel Equalization Techniques for Speech Dereverberation“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Kyoto, Japan, März 2012, S. 537–540. DOI: 10.1109/ICASSP.2012.6287935.

[91] Ina Kodrasi und Simon Doclo. „The Effect of Inverse Filter Length on the Robustness of Acoustic Multichannel Equalization“. In: *Proceedings of the European Signal Processing Conference*. Bucharest, Romania, Aug. 2012, S. 2442–2446.

[92] Ina Kodrasi, Timo Gerkmann und Simon Doclo. „Frequency-Domain Single-Channel Inverse Filtering for Speech Dereverberation: Theory and Practice“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Florence, Italy, Mai 2014, S. 5214–5218.

[93] Ina Kodrasi, Stefan Goetze und Simon Doclo. „A Perceptually Constrained Channel Shortening Technique for Speech Dereverberation“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vancouver, BC, Canada, Mai 2013, S. 151–155. DOI: 10.1109/ICASSP.2013.6637627.

[94] Ina Kodrasi, Stefan Goetze und Simon Doclo. „Increasing the Robustness of Acoustic Multichannel Equalization by Means of Regularization“. In: *Proceedings of the IEEE International Workshop on Acoustic Signal Enhancement*. Aachen, Germany, Sep. 2012.

[95] Ina Kodrasi, Stefan Goetze und Simon Doclo. „Regularization for Partial Multichannel Equalization for Speech Dereverberation“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 21.9 (Sep. 2013), S. 1879–1890. DOI: 10.1109/TASL.2013.2260743.

[96] Mihailo Kolundžija, Christof Faller und Martin Vetterli. „Multi-Channel Low Frequency Room Equalization Using Perceptually Motivated Constrained Optimization“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Kyoto, Japan, März 2012, S. 533–536. DOI: 10.1109/ICASSP.2012.6287934.

[97] R. Kürer. „Zur Gewinnung von Einzahlkriterien bei Impulsmessung in der Raumakustik“. In: *Acustica* 21 (1969), S. 370–372.

[98] Heinrich Kuttruff. *Room Acoustics*. Spon Press, 2009.

[99] K. Lebart, J. M Boucher und P. N. Denbigh. „A New Method Based on Spectral Subtraction for Speech Dereverberation“. In: *Acta Acustica united with Acustica* 87 (2001), S. 359–366.

[100] Tobias Lentz u. a. „Virtual Reality System with Integrated Sound Field Simulation and Reproduction“. In: *EURASIP Journal on Applied Signal Processing* 2007.1 (Jan. 2007), S. 187–205. DOI: 10.1155/2007/70540.

[101] Felicia Lim und Patrick A. Naylor. „Relaxed Multichannel Least Squares with Constrained Initial Taps for Multichannel Dereverberation“. In: *Proceedings of the IEEE International Workshop on Acoustic Signal Enhancement*. Aachen, Germany, Sep. 2012.

[102] Philipos C. Loizou. *Speech Enhancement: Theory and Practice*. CRC Press, Taylor und Francis Group, 2007.

[103] David Lubman. „Fluctuations of Sound with Position in a Reverberant Room“. In: *Journal of the Acoustical Society of America* 44.6 (1968), S. 1491–1502. DOI: 10.1121/1.1911287.

[104] Richard F. Lyon, Andreas G. Katsiamis und Emmanuel M. Drakakis. „History and Future of Auditory Filter Models“. In: *Proceedings of IEEE Internation Symposium on Circuits and Systems*. Paris, France, Mai 2010, S. 3809–3812. DOI: 10.1109/ISCAS.2010.5537724.

[105] John I. Makhoul. „Linear Prediction: A Tutorial Review“. In: *Proceedings of the IEEE* 63.4 (Apr. 1975), S. 561–580. DOI: 10.1109/PROC.1975.9792.

[106] John I. Makhoul und Jared J. Wolf. *Linear Prediction and the Spectral Analysis of Speech*. Techn. Ber. 2304. Cambridge, Massachusetts: Bolt, Beranek und Newman Inc., Aug. 1972.

[107] Shoji Makino, Yutaka Kaneda und Nobuo Koizumi. „Exponentially Weighted Stepsize NLMS Adaptive Filter Based on the Statistics of a Room Impulse Response“. In: *IEEE Transactions on Speech and Audio Processing* 1.1 (Jan. 1993), S. 101–108. DOI: 10.1109/89.221372.

[108] Richard K. Martin u. a. „Unification and Evaluation of Equalization Structures and Design Algorithms for Discrete Multitone Modulation Systems“. In: *IEEE Transactions on Signal Processing* 53.10 (Okt. 2005), S. 3880–3894. DOI: 10.1109/TSP.2005.855432.

[109] Radoslaw Mazur. „Neue Ansätze zur Lösung des Skalierungs- und Permutationsproblems in konvolutiver blinder Quellentrennung“. Diss. Universität zu Lübeck, 2010.

[110] Radoslaw Mazur, Jan Ole Jungmann und Alfred Mertins. „On CUDA Implementation of a Multichannel Room Impulse Response Reshaping Algorithm Based on p-Norm Optimization“. In: *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. New Paltz, New York, USA, Okt. 2011, S. 305–308. DOI: 10.1109/ASPAA.2011.6082310.

[111] Radoslaw Mazur, Jan Ole Jungmann und Alfred Mertins. „Optimized Gradient Calculation for Room Impulse Response Reshaping Algorithm Based on p-Norm Optimization“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Kyoto, Japan, März 2012, S. 185–188. DOI: 10.1109/ICASSP.2012.6287848.

[112] Radoslaw Mazur und Alfred Mertins. „An Approach for Solving the Permutation Problem of Convolutive Blind Source Separation Based on Statistical Signal Models“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 17.1 (Jan. 2009), S. 117–126. DOI: 10.1109/TASL.2008.2005349.

[113] Sanjeev Mehrotra, Wei ge Chen und Zhengyou Zhang. „Interpolation of Combined Head and Room Impulse Response for Audio Spatialization“. In: *Proc. 13th IEEE International Workshop on Multimedia Signal Processing (MMSP)*. IEEE. Hangzhou, China, Okt. 2011, S. 1–6.

[114] Tiemin Mei und Alfred Mertins. „On the Robustness of Room Impulse Response Reshaping“. In: *Proceedings of the IEEE International Workshop on Acoustic Echo and Noise Control*. Tel Aviv, Israel, Aug. 2010.

[115] Tiemin Mei, Alfred Mertins und Markus Kallinger. „Room Impulse Response Reshaping/Shortening Based on Least Mean Squares Optimization With Infinity Norm Constraint“. In: *Proc. 16th Int. Digital Signal Processing Conf. (DSP 2009)*. 2009, S. 1–6. DOI: 10.1109/ICDSP.2009.5201237.

[116] Tiemin Mei, Alfred Mertins und Markus Kallinger. „Room Impulse Response Shortening with Infinity-Norm Optimization“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Taipei, Taiwan, Apr. 2009, S. 3745–3748. DOI: 10.1109/ICASSP.2009.4960441.

[117] Andreas Meister. *Numerik linearer Gleichungssysteme*. 2. Aufl. Vieweg+Teubner, 2005.

[118] Peter J. W. Melsa, Richard C. Younce und Charles E. Rohrs. „Impulse Response Shortening for Discrete Multitone Transceivers“. In: *IEEE Transactions on Communications* 44.12 (Dez. 1996), S. 1662–1672. DOI: 10.1109/26.545896.

[119] Alfred Mertins. *Signaltheorie*. 3. Aufl. Springer Vieweg, 2013. DOI: 10.1007/978-3-8348-8109-0.

[120] Alfred Mertins, Tiemin Mei und Markus Kallinger. „Room Impulse Response Shortening/Reshaping With Infinity- and p-Norm Optimization“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 18.2 (Feb. 2010), S. 249–259. DOI: 10.1109/TASL.2009.2025789.

[121] Erwin Meyer und Rolf Thiele. „Raumakustische Untersuchung in zahlreichen Konzertsälen und Rundfunkstudios unter Anwendung neuer Meßverfahren“. In: *Acustica* 6 (1956), S. 425.

[122] Rémi Mignot, Gilles Chardon und Laurent Daudet. „Compressively Sampling the Plenacoustic Function“. In: *Proc SPIE XIV. Conf. Wavelets and Sparsity*. San Diego, CA, USA, Sep. 2011. DOI: 10.1117/12.892761.

[123] Rémi Mignot, Laurent Daudet und François Ollivier. „Compressed Sensing for Acoustic Response Reconstruction: Interpolation of the Early Part“. In: *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. New Paltz, New York, USA, Okt. 2011, S. 225–228.

[124] Rémi Mignot, Laurent Daudet und François Ollivier. „Room Reverberation Reconstruction: Interpolation of the Early Part Using Compressed Sensing“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 21.11 (Nov. 2013), S. 2301–2312. DOI: 10.1109/TASL.2013.2273662.

[125] Masato Miyoshi und Yutaka Kaneda. „Inverse Filtering of Room Acoustics“. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 36.2 (Feb. 1988), S. 145–152. DOI: 10.1109/29.1509.

[126] Dennis R. Morgan, Jacob Benesty und Man Mohan Sondhi. „On the Evaluation of Estimated Impulse Responses“. In: *IEEE Signal Processing Letters* 5.7 (Juli 1998), S. 174–176. DOI: 10.1109/97.700920.

[127] John N. Mourjopoulos. „Digital Equalization of Room Acoustics“. In: *Journal of the Audio Engineering Society* 42.11 (Nov. 1994), S. 884–900.

[128] John N. Mourjopoulos. „On the Variation and Invertibility of Room Impulse Response Functions“. In: *Journal of Sound and Vibration* 102.2 (Sep. 1985), S. 217–228. DOI: 10.1016/S0022-460X(85)80054-7.

[129] Tomohiro Nakatani und Masato Miyoshi. „Blind Dereverberation of Single Channel Speech Signal Based on Harmonic Structure“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Hong Kong, China, Apr. 2003, S. 92–95. DOI: 10.1109/ICASSP.2003.1198724.

[130] Stephen T. Neely und Jont B. Allen. „Invertibility of a Room Impulse Response“. In: *Journal of the Acoustical Society of America* 66.1 (Juli 1979), S. 165–169. DOI: 10.1121/1.383069.

[131] Philip A. Nelson, Felipe Orduña-Bustamante und Hareo Hamada. „Inverse Filter Design an Equalization Zones in Multichannel Sound Reproduction“. In: *IEEE Transactions on Speech and Audio Processing* 3.3 (Mai 1995), S. 185–192. DOI: 10.1109/89.388144.

[132] Alan V. Oppenheim und Ronald W. Schafer. *Discrete-Time Signal Processing*. Prentice Hall Inc., 1999.

[133] Jukka Pätynen, Sakari Tervo und Tapio Lokki. „Analysis of Concert Hall Acoustics via Visualizations of Time-Frequency and Spatiotemporal Responses“. In: *Journal of the Acoustical Society of America* 133.2 (Feb. 2013), S. 842–857. DOI: 10.1121/1.4770260.

[134] Roger Penrose. „A Generalized Inverse for Matrices“. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 51.3 (Juli 1955), S. 406–413. DOI: 10.1017/s0305004100030401.

[135] Jean-Dominique Polack. „La transmission de l'énergie sonore dans les salles“. Diss. Le Mans, France: Université du Maine, 1988.

[136] William H. Press u. a. *Numerical Recipes in C++*. 2. Aufl. Cambridge University Press, 2002.

[137] David H. Raab. „Forward and Backward Masking between Acoustic Clicks“. In: *Journal of the Acoustical Society of America* 33.2 (Feb. 1961), S. 137–139. DOI: 10.1121/1.1908600.

[138] Lawrence Rabiner, James H. McClellan und Thomas W. Parks. „FIR Digital Filter Design Techniques Using Weighted Chebyshev Approximation“. In: *Proceedings of the IEEE* 63.4 (1975), S. 595–610. DOI: 10.1109/PROC.1975.9794.

[139] Biljana D. Radlović, R. C. Williamson und Rodney A. Kennedy. „Equalization in an Acoustic Reverberant Environment: Robustness Results“. In: *IEEE Transactions on Speech and Audio Processing* 8.3 (Mai 2000), S. 311–319. DOI: 10.1109/89.841213.

[140] Biljana D. Radlović, R. C. Williamson und Rodney A. Kennedy. „On the Poor Robustness of Sound Equalization in Reverberant Environments“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Phoenix, Arizona, USA, Mai 1999, S. 881–884. DOI: 10.1109/ICASSP.1999.759812.

[141] Walter Reichardt, O. Abdel Alim und W. Schmidt. „Abhängigkeit der Grenzen zwischen brauchbarer und unbrauchbarer Durchsichtigkeit von der Art des Musikmotivs, der Nachhallzeit und der Nachhalleinsatzzeit“. In: *Applied Acoustics* 7 (1974), S. 243.

[142] Martin Rothbucher u. a. „Comparison of Head-Related Impulse Response Measurement Approaches“. In: *Journal of the Acoustical Society of America* 134.2 (Aug. 2013), EL223–EL229. DOI: 10.1121/1.4813592.

[143] Jason Sanders und Edward Kandrot. *CUDA by Example: An Introduction to General-Purpose GPU Programming*. 1. Aufl. Addison-Wesley Professional, 2010.

[144] H. Schneider. „Zur Messung der Sprachverständlichkeit“. In: *Frequenz* 9.6 (Juni 1955), S. 199–205. DOI: 10.1515/FREQ.1955.9.6.199.

[145] Manfred R. Schroeder. „Frequency-Correlation Functions of Frequency Responses in Rooms“. In: *Journal of the Acoustical Society of America* 34.12 (Dez. 1962), S. 1819–1823. DOI: 10.1121/1.1909136.

[146] Manfred R. Schroeder. „Integrated-Impulse Method Measuring Sound Decay Without Using Impulses“. In: *Journal of the Acoustical Society of America* 66.2 (Aug. 1979), S. 497–500. DOI: 10.1121/1.383103.

[147] Manfred R. Schroeder. „New Method of Measuring Reverberation Time“. In: *Journal of the Acoustical Society of America* 37.3 (1965), S. 409–412. DOI: 10.1121/1.1909343.

[148] Manfred R. Schroeder und Heinrich Kuttruff. „On Frequency Response Curves in Rooms - Comparison of Experimental, Theoretical and Monte Carlo Results for the Average Frequency Spacing Between Maxima“. In: *Journal of the Acoustical Society of America* 34.1 (Jan. 1962), S. 76–80. DOI: 10.1121/1.1909022.

[149] Claude Elwood Shannon. „Communication in the Presence of Noise“. In: *Proceedings of the IRE* 37.1 (Jan. 1949), S. 10–21.

[150] Myung-Suk Song u. a. „An Interactive 3-D Audio System With Loudspeakers“. In: *IEEE Transactions on Multimedia* 13.5 (Okt. 2011), S. 844–855. DOI: 10.1109/TMM.2011.2162581.

[151] Myung-Suk Song u. a. „Enhancing Loudspeaker-Based 3D Audio with Room Modeling“. In: *Proc. 2010 IEEE International Workshop on Multimedia Signal Processing (MMSP 2010)*. Okt. 2010, S. 34–39. DOI: 10.1109/MMSP.2010.5661990.

[152] Myung-Suk Song u. a. „Personal 3D Audio System with Loudspeakers“. In: *Proc. 2010 IEEE International Conference on Multimedia and Expo (ICME 2010)*. Singapur, Juli 2010, S. 1600–1605. DOI: 10.1109/ICME.2010.5583184.

[153] Jan-Jakob Sonke. „Variable Acoustics by Wave Field Synthesis“. Diss. Delft, The Netherlands: Technische Universiteit Delft, 2000.

[154] Sascha Spors. „Schallfeldsynthese: Stand der Technik und Perspektiven“. In: *Proceedings of the German Annual Conference on Acoustics (DAGA)*. Duesseldorf, Germany, März 2011, S. 7–8.

[155] Sascha Spors, Achim Kuntz und Rudolf Rabenstein. „An Approach to Listening Room Compensation with Wave Field Synthesis“. In: *24th International Conference on Multichannel Audio, Audio Engineering Society*. Banff, Canada, Juni 2003, S. 49–52.

[156] Sasche Spors und Jens Ahrens. „A Comparison of Wave Field Synthesis and Higher-Order Ambisonics with Respect to Physical Properties and Spatial Sampling“. In: *Proc. 125th Convention of the Audio Engineering Society*. San Francisco, CA, USA, Okt. 2008.

[157] Guy-Bart Stan, Jean-Jacques Embrechts und Dominique Archambeau. „Comparison of Different Impulse Response Measurement Techniques“. In: *Journal of the Audio Engineering Society* 50.4 (2002), S. 249–262.

[158] Rebecca Stewart und Mark Sandler. „Statistical Measures of Early Reflections of Room Impulse Responses“. In: *Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07)*. Bordeaux, France, Sep. 2007.

[159] Stefan Strahl und Alfred Mertins. „Analysis and Design of Gammatone Signal Models“. In: *Journal of the Acoustical Society of America* 126.5 (Nov. 2009), S. 2379–2389. DOI: 10.1121/1.3212919.

[160] Ingo R. Titze. *Principles of Voice Production*. Allyn & Bacon, 1994.

[161] Mahdi Triki und Dirk T. M. Slock. „Iterated Delay and Predict Equalization for Blind Speech Dereverberation“. In: *Proceedings of the IEEE International Workshop on Acoustic Echo and Noise Control*. Paris, France, Sep. 2006.

[162] John Vanderkooy. „Aspects of MLS Measuring Systems“. In: *Journal of the Audio Engineering Society* 42.4 (Apr. 1994), S. 219–231.

[163] Bo Wang u. a. „Generalized Channel Impulse Response Shortening for Discrete Multitone Transceivers“. In: *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove, CA, USA, Okt. 1999, S. 276–280. DOI: 10.1109/ACSSC.1999.832336.

[164] Darren B. Ward. „Joint Least Squares Optimization for Robust Acoustic Crosstalk Cancellation“. In: *IEEE Transactions on Speech and Audio Processing* 8.2 (Feb. 2000), S. 211–215.

[165] Darren B. Ward. „On the Performance of Acoustic Crosstalk Cancellation in a Reverberant Environment:“ In: *Journal of the Acoustical Society of America* 110.2 (Aug. 2001), S. 1195–1198. DOI: 10.1121/1.1386635.

[166] Anna Warzybok u. a. „Subjective Speech Quality and Speech Intelligibility Evaluation of Single-Channel Dereverberation Algorithms“. In: *Proceedings of the IEEE International Workshop on Acoustic Signal Enhancement*. Antibes, France, Sep. 2014.

[167] Richard V. Waterhouse. „Statistical Properties of Reverberant Sound Fields“. In: *Journal of the Acoustical Society of America* 43.6 (1968), S. 1436–1444. DOI: 10.1121/1.1911004.

[168] Wancheng Zhang, Nikolay D. Gaubitch und Patrick A. Naylor. „Computationally Efficient Equalization of Room Impulse Responses Robust to System Estimation Errors“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Las Vegas, NV, USA, Apr. 2008, S. 4025–4028. DOI: 10.1109/ICASSP.2008.4518537.

[169] Wancheng Zhang, Emanuël A. P. Habets und Patrick A. Naylor. „A System-Identification-Error-Robust Method for Equalization of Multichannel Acoustic Systems“. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Dallas, Texas, USA, März 2010, S. 109–112. DOI: 10.1109/ICASSP.2010.5496153.

[170] Wancheng Zhang, Emanuël A. P. Habets und Patrick A. Naylor. „On the Use of Channel Shortening in Multichannel Acoustic System Equalization“. In: *Proceedings of the IEEE International Workshop on Acoustic Echo and Noise Control*. Tel Aviv, Israel, Aug. 2010.

[171] Wancheng Zhang, Andy W. H. Khong und Patrick A. Naylor. „Adaptive Inverse Filtering of Room Acoustics“. In: *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove, CA, USA, Okt. 2008, S. 26–29.

[172] Eberhard Zwicker und Hugo Fastl. *Psychoacoustics: Facts and Models*. 3rd. Springer, 2007.

[173] Marina Dana Ţopa u. a. „Experimental Acoustic Evaluation of an Auditorium“. In: *Advances in Acoustics and Vibration* 2012 (2012), S. 8. DOI: 10.1155/2012/868247.