

Rückkopplungsunterdrückung für Innenraumkommunikationssysteme

Dissertation

zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften
(Dr.-Ing.)
der Technischen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Philipp Bulling

Ulm 2018

1. Berichterstatter: Prof. Dr.-Ing. Gerhard Schmidt
2. Berichterstatter: Prof. Dr.-Ing. Jürgen Freudenberger

Datum der mündlichen Prüfung: 10. September 2018

Danksagung

Diese Dissertation entstand während meiner Arbeit bei der Daimler AG in Ulm. Ermöglicht wurde sie durch Dr.-Ing. Klaus Linhard, der die Arbeit vor Ort betreut hat. Hierfür gebührt ihm mein größtmöglicher Dank. Für viele Themen, die ich im Rahmen dieser Arbeit untersucht habe, hat er den Grundstein gelegt und er ist mir auch darüber hinaus stets mit Rat und Tat zur Seite gestanden.

Nicht weniger dankbar bin ich auch meinem Doktorvater Prof. Dr.-Ing. Gerhard Schmidt von der Christian-Albrechts-Universität zu Kiel, durch den ich trotz der großen Entfernung stets eine hervorragende Betreuung erfahren habe. Zu dieser Betreuung zählen unter anderem die mehrwöchigen Aufenthalte in Kiel sowie die regelmäßigen telefonischen Besprechungen und die ausführliche Beantwortung meiner Zwischenberichte.

Weiter danke ich meinem zweiten Gutachter Prof. Dr.-Ing. Jürgen Freudenberger für die Übernahme des Koreferates und die damit verbundenen Bemühungen. Seine hilfreichen Anmerkungen haben die Arbeit letztlich abgerundet.

Die Prüfungskommission setzte sich darüber hinaus aus Prof. Dr.-Ing. habil. Thomas Meurer, Prof. Dr.-Ing. Michael Höft sowie dem Vorsitzenden Prof. Dr.-Ing. Jeffrey McCord zusammen. Ihnen sei an dieser Stelle ebenfalls gedankt.

Ich danke auch meinen Kollegen bei Daimler am Standort Ulm. Insbesondere gilt mein Dank meinen Teamkollegen Arthur Wolf und Harald Schnepf für ihre Unterstützung und die freundschaftliche Zusammenarbeit. Dass ich dank ihrer Hilfe uneingeschränkt Zugriff auf Versuchsfahrzeuge, Labore und Messeinrichtungen hatte, hat mir die Arbeit enorm erleichtert. An dieser Stelle sei auch unseren studentischen Mitarbeitern gedankt, die alle ihren Teil zu dieser Arbeit beigetragen haben. In chronologischer Reihenfolge waren dies: Gordon Axnick, Ranjith Ponnusamy, Lorik Bajrami, Bastian Biedermann, Tobias Klawonn, Florian Unseld, Tobias Welz, Tim Benedikt Kupke und Florian Aue.

Den Mitarbeitern, Doktoranden und mittlerweile Doktoren vom Lehrstuhl für digitale Signalverarbeitung und Systemtheorie der Christian-Albrechts-Universität zu Kiel danke ich dafür, dass ich während meiner Forschungsaufenthalte in Kiel stets freundlich aufgenommen wurde und immer für das passende Rahmenprogramm gesorgt war. Hervorheben möchte ich Marco Gimm für die gute Zusammenarbeit bei unseren gemeinsamen Veröffentlichungen.

Besonderen Dank richte ich an dieser Stelle an meine Eltern Anne und Dieter Bulling sowie an meine Schwester Jessica Bulling für ihren Rückhalt und ihre unermüdliche Unterstützung. Ihr moralischer Beistand war mir immer die größte Hilfe.

Ulm, den 14. September 2018

Philipp Bulling

Kurzfassung

Da die Kommunikation zwischen den Insassen eines Kraftfahrzeugs aufgrund hoher Hintergrundgeräuschpegel oftmals erschwert ist, werden in verschiedenen modernen Fahrzeugen sog. Innenraumkommunikationssysteme eingesetzt. Diese verbessern die Kommunikation, indem sie die Stimmen der Sprecher mit Mikrofonen aufzeichnen und über Lautsprecher bei den Zuhörern wiedergeben. Die Herausforderung bei diesen Systemen ist die elektro-akustische Rückkopplung. Diese wird dadurch verursacht, dass das Mikrofon nicht nur das lokale Sprachsignal, sondern auch das Lautsprechersignal aufzeichnet. Ohne Gegenmaßnahmen resultiert die Rückkopplung in einem unangenehmen, lauten Pfeifen.

Die Problematik der elektro-akustischen Rückkopplung ist für Innenraumkommunikationssysteme noch nicht umfassend gelöst. Deshalb werden in der vorliegenden Arbeit Verfahren vorgeschlagen, welche in der Lage sind, die Rückkopplungen mittels digitaler Signalverarbeitung zu unterdrücken. Der Hauptteil dieser Arbeit beschäftigt sich mit der adaptiven Rückkopplungskompensation. Bei dieser wird versucht, die Impulsantwort zwischen den Lautsprechern und dem Mikrofon mit einem adaptiven Filter zu schätzen. Die Schwierigkeit ist hierbei die starke Korrelation zwischen Lautsprechersignal und lokaler Sprache, welche verhindert, dass das Filter zur korrekten Lösung konvergieren kann. Um die Konvergenz zu verbessern, wird eine neuartige Schrittweitensteuerung vorgestellt. Diese nutzt zur Aktualisierung der Filterkoeffizienten kurze Zeitpunkte im Nachhall, während derer die Signale unkorreliert sind. Ergänzend zur Rückkopplungskompensation wird anschließend ein Postfilter vorgestellt. Dessen Aufgabe ist es, diejenigen Rückkopplungsanteile, welche nach der Rückkopplungskompensation noch vorhanden sind, mittels eines Wiener-Filters zu unterdrücken. Hierfür wird das Postfilter abhängig vom Abgleichzustand des adaptiven Filters gesteuert. Abschließend werden zwei Verfahren vorgestellt, welche die Sprachqualität des Systems verbessern. Zum einen wird die Verbesserung des Klangs des Systems durch eine automatische Entzerrung beschrieben. Zum anderen wird gezeigt, dass die Sprachverständlichkeit eines Signals durch gezieltes Hinzufügen von Harmonischen verbessert werden kann.

Bei den theoretischen Untersuchungen wird in dieser Arbeit auch die praktische Umsetzbarkeit der Algorithmen berücksichtigt. Zu diesem Zweck werden die Verfahren in ein eigens entwickeltes Echtzeit-Rahmenwerk integriert und während zahlreicher Testfahrten mit Demonstratorfahrzeugen im realen Betrieb erprobt. Bei diesen praktischen Tests zeigt sich ein deutlicher Gewinn hinsichtlich Stabilität und Sprachqualität im Vergleich zu bisher bekannten Verfahren.

Abstract

The communication between the passengers inside a car can be difficult due to large background noise levels. It can be improved with so-called in-car communication systems. These systems capture the voice of talkers by means of microphones and play it back via loudspeakers close to the listeners. However, the challenge is the electro-acoustic feedback, which occurs when the microphone not only captures the local speech but also the loudspeaker signal. Without countermeasures, this feedback results in annoying howling sounds.

The problem of the electro-acoustic feedback has not yet been solved for in-car communication systems. Therefore, in this work techniques to suppress the feedback by means of digital signal processing are presented. The main part of this work focuses on adaptive feedback cancellation. Here, the impulse response between loudspeaker and microphone is estimated with an adaptive filter. The difficulty is a strong correlation between loudspeaker and local speech that prevents the adaptive filter from converging towards the desired solution. In order to improve convergence, a novel stepsize control is presented. As signals are not correlated during reverberation, the stepsize control exploits reverberant signal periods to update the filter coefficients. In addition to the adaptive feedback canceler, a postfilter is presented. The task of the postfilter is to suppress the residual feedback that remains after the feedback cancellation, by means of a Wiener-filter. Therefore, the postfilter is controlled depending on the adaptive filter's state of convergence. Finally, two techniques to improve the speech quality are presented. Firstly, an automatic equalizer is described that improves the sound quality. Secondly, it is shown that speech intelligibility can be improved by adding harmonics to a speech signal.

Besides the theoretical investigations, in this work also the practical realization of the algorithms is regarded. Therefore, the algorithms are integrated into a specially developed real-time framework and tested in demonstration cars under realistic conditions during numerous test drives. These test drives show a significant increase of both stability and speech quality compared to existing approaches.

Inhaltsverzeichnis

Abbildungsverzeichnis	xv
Tabellenverzeichnis	xvii
Abkürzungen und Notation	xix
1 Einleitung	1
1.1 Innenraumkommunikationssysteme	1
1.2 Inhalt dieser Arbeit	3
2 Randbedingungen im Fahrzeug	5
2.1 Demonstratorfahrzeug	5
2.2 Messung der Impulsantworten im Fahrzeug	6
2.2.1 Methode	8
2.2.2 Ergebnisse	10
2.3 Anforderungen an die Signalverarbeitung	15
2.3.1 Automatische Anpassung	16
2.3.2 Geringe Laufzeit	17
2.3.3 Unterdrückung der Hintergrundgeräusche	17
2.4 Zusammenfassung	20
3 Grundlagen der Rückkopplungsunterdrückung	21
3.1 Stabilität rückgekoppelter Systeme	21
3.2 Stand der Technik	23
3.3 Prinzip der adaptiven Rückkopplungskompensation	26
3.3.1 Theorie adaptiver Filter	27
3.3.2 Implementierung adaptiver Filter im Frequenzbereich	31
3.4 Prinzip der Rückkopplungsreduktion mit Wiener-Filter	36
3.4.1 Herleitung des Wiener-Filters	36
3.4.2 Gewichtete <i>Overlap-Add</i> -Filterbank	38
3.5 Verwendete Filterbankstruktur	39
3.5.1 Kombination der Filterbänke	41
3.5.2 Parametrierung	43
3.6 Zusammenfassung	44
4 Schrittweitensteuerung für die adaptive Rückkopplungskompensation	47
4.1 Optimale Schrittweite	47
4.1.1 Herleitung der optimalen Schrittweite	48
4.1.2 Näherungen und zu schätzende Größen	49

4.2	Nachhall-basierte Schrittweitensteuerung	50
4.2.1	Nachhall	50
4.2.2	Detektion des Nachhalls	52
4.3	Verstärkungsregelung	56
4.4	Schätzen des Systemabstands	57
4.4.1	Nullen am Ende der Impulsantwort	57
4.4.2	Verfahren mit Minimumschätzer	59
4.5	Detektion von Raumänderungen	62
4.5.1	Schattenfilter	62
4.5.2	Anstieg des Fehlersignals	63
4.5.3	Vergleich der hochpass- und tiefpassgefilterten Fehlerleistung . . .	65
4.5.4	Vergleich der mittleren Leistungen von Mikrofonsignal und Fehler- signal	66
4.6	Verfahren zur Detektion von Aufschwingen	69
4.7	Ergebnisse und Simulationen	70
4.7.1	Simulation mit konstanter Impulsantwort	71
4.7.2	Simulation von Raumänderungen	75
4.8	Zusammenfassung	78
5	Postfilter zur Reduktion der Restrückkopplung	81
5.1	Modellbasierte Rückkopplungsreduktion	81
5.2	Kombination von Rückkopplungskompensation und Rückkopplungsreduk- tion	86
5.3	Varianten des Hallmodells	87
5.3.1	Variante A	88
5.3.2	Variante B	88
5.4	Ergebnisse und Simulationen	89
5.4.1	Objektive Evaluierung	90
5.4.2	Subjektive Evaluierung	94
5.5	Zusammenfassung	96
6	Verfahren zur Verbesserung der Sprachqualität	97
6.1	Automatische Entzerrung	97
6.1.1	Adaptives Filter zum Schätzen des kurzen Pfades	98
6.1.2	Zielfrequenzgang des Entzerrfilters	100
6.1.3	Ergebnis	102
6.1.4	Zusammenfassung	104
6.2	Verbesserung der Sprachverständlichkeit	105
6.2.1	Prinzip des <i>Exciters</i>	105
6.2.2	Parametrierung des Algorithmus	107
6.2.3	Ergebnis und Evaluierung	109
6.2.4	Zusammenfassung	113

7 Zusammenfassung und Ausblick	115
7.1 Zusammenfassung	115
7.2 Ausblick	116
A Herleitung der optimalen Schrittweite	119
B Echtzeit-Rahmenwerk RACE	123
Literaturverzeichnis	127

Abbildungsverzeichnis

1.1	Block-Diagramm eines ICC-Systems.	2
2.1	Das Demonstratorfahrzeug (Mercedes-Benz Viano, Baureihe 639).	6
2.2	Lautsprecher-/Mikrofonpositionen im Demonstratorfahrzeug.	7
2.3	Ein Kunstkopf im Demonstratorfahrzeug mit den verbauten Mikrofonen und Lautsprechern.	8
2.4	Der Sinussweep für die Impulsantwortmessung.	9
2.5	Schematischer Ablauf der Impulsantwortmessung.	9
2.6	Der Frequenzgang der verwendeten Mikrofone.	10
2.7	Die Frequenzgänge von den hinteren Lautsprechern auf die vorderen Mi- krofone und von den vorderen Lautsprechern auf die hinteren Mikrofone.	11
2.8	Energieabklingkurven im leeren Fahrzeug.	12
2.9	Die Frequenzgänge und zugehörigen Impulsantworten der hinteren Pfade.	13
2.10	Vergleich der Impulsantworten von den hinteren Lautsprechern zum Fahr- ermikrofon bei leerem und voll besetztem Fahrzeug.	14
2.11	Die Frequenzgänge, wenn der Passagier sein Knie vor den Lautsprecher hält.	15
2.12	Die Frequenzgänge, wenn der Fahrer mit seiner Hand das Mikrofon verdeckt.	15
2.13	Block-Diagramm eines bidirektionalen ICC-Systems.	16
2.14	Leistungsdichtespektren verschiedener Fahrgeräusche.	18
2.15	Leistungsdichtespektren von Sprache und Fahrgeräusch.	19
3.1	Schematisches Blockdiagramm einer geschlossenen elektro-akustischen Schlei- fe.	21
3.2	Schematisches Blockdiagramm eines rückgekoppelten Systems mit adap- tiver Rückkopplungskompensation.	26
3.3	Adaptives Filter bei der Systemidentifikation.	27
3.4	Schnelle Faltung mit einer <i>Overlap-Save</i> -Filterbank.	32
3.5	Zusammensetzen des Ausgangssignals bei der <i>Overlap-Save</i> -Methode.	34
3.6	Blockdiagramm der Implementierung eines adaptiven Filters im Teilband.	34
3.7	Wiener-Filter im Zeitbereich zur Reduktion der Rückkopplung.	36
3.8	Blockweise Verarbeitung bei der gewichteten <i>Overlap-Add</i> -Methode.	40
3.9	Blockdiagramm der in dieser Arbeit verwendeten Filterbankstruktur.	42
4.1	Blockdiagramm der adaptiven Rückkopplungskompensation.	48
4.2	Der Nachhall in einer geschlossenen elektro-akustischen Schleife.	52
4.3	Das Prinzip der Nachhalldetektion.	53
4.4	Verläufe der Größen für die Nachhalldetektion.	55

4.5	Zerlegung der zu schätzenden Impulsantwort in Teilfilter.	58
4.6	Kennlinie des Minimumschätzers.	61
4.7	Schattenfilter zur Detektion von Raumänderungen.	63
4.8	Kennlinie des Minimumschätzers mit zusätzlicher Schwelle zur Detektion von Raumänderungen.	64
4.9	Blockdiagramm mit allen beteiligten Teilsystemen.	71
4.10	Der zeitliche Verlauf der Schrittweite.	73
4.11	Die Spektren von Schrittweite und lokaler Sprache während Sprachaktivität.	73
4.12	Der zeitliche Verlauf der mittleren Leistungen von Lautsprecher- und Fehler- signal, sowie des Verhältnisses $EXR(\mu, k)$ im Teilband $\mu = 4$	74
4.13	Der zeitliche Verlauf der Steuergröße des Minimumschätzers $\beta_{\min}(\mu, k)$ im Teilband $\mu = 4$	75
4.14	Der zeitliche Verlauf von realem und geschätztem Systemabstand im Teil- band $\mu = 4$, sowie der Verlauf der Verstärkung.	76
4.15	Der zeitliche Verlauf der für die Raumänderungsdetektion benötigten Grö- ßen.	77
4.16	Der zeitliche Verlauf von geschätztem Systemabstand, der für die Schät- zung benötigten Größen im Teilband $\mu = 35$ sowie der Verstärkung.	78
4.17	Der zeitliche Verlauf der quadrierten Betragsfrequenzgänge von realer Im- pulsantwort und adaptivem Filter.	79
5.1	Blockdiagramm der modellbasierten Rückkopplungsreduktion mit Wiener- Filter.	82
5.2	Die Kopplungsfaktoren für das Postfilter zur Rückkopplungsreduktion.	84
5.3	Schematische Darstellung des rekursiven Hallmodells.	85
5.4	Spektrogramme des modellierten und des gemessenen Übertragungsver- haltens zwischen den hinteren Lautsprechern und dem Fahrermikrofon.	85
5.5	Zusammensetzen der geschätzten Impulsantwort aus den Teilfiltern des Rückkopplungskompensators.	86
5.6	Das Abklingverhalten des rekursiven Hallmodells unter der idealen Be- dingung, dass der Kompensator in den ersten M Blöcken perfekt abgleicht.	87
5.7	Variante A des Hallmodells.	88
5.8	Variante B des Hallmodells.	89
5.9	Blockdiagramm der Teilsysteme.	90
5.10	Ausschnitt des zeitlichen Verlaufs der Masken zur Evaluierung des Post- filters.	92
5.11	Schematisches Blockdiagramm zur Bestimmung der mittleren Leistungen.	93
5.12	Ergebnis des Hörtests zur Evaluierung der Postfiltervarianten.	95
6.1	Blockdiagramm eines ICC-Systems mit automatischer Entzerrung des Zu- hörerpfades.	99
6.2	Schematische Darstellung der Frequenzbereichsimplementierung des ad- aptiven Filters zur Schätzung des kurzen Pfades.	99
6.3	Der Betragsfrequenzgang des konvergierten adaptiven Filters.	100

6.4	Betragsfrequenzgänge zur Berechnung des Entzerrfilters.	102
6.5	Das Konvergenzverhalten des adaptiven Filters für die automatische Entzerrung.	103
6.6	Der zeitliche Verlauf des Entzerrfilters als Spektrogramm.	103
6.7	Der Frequenzgang am Zuhörerplatz mit und ohne Entzerrung nach 19,5 s Simulationszeit.	104
6.8	Prinzipielle Verarbeitung beim <i>Exciter</i>	106
6.9	Blockschaltbild des <i>Exciters</i>	106
6.10	Der für weißes Eingangsruschen am Ausgang gemessene Frequenzgang des <i>Exciters</i>	108
6.11	Die Kennlinien des <i>Exciters</i>	109
6.12	Die Klirrfaktoren der Harmonischen.	110
6.13	Spektrogramme einer weiblichen Sprecherin aus dem Szenario S4.	111
6.14	Das Ergebnis der drei Szenarien zur Evaluierung der Sprachqualität.	112
6.15	Das Ergebnis der Evaluierung des Szenarios S4.	113
B.1	Die grafische Benutzeroberfläche von RACE.	123
B.2	Die grafische Benutzeroberfläche des Moduls für die Rückkopplungskompensation.	124

Tabellenverzeichnis

3.1	Vor-/Nachteile der Filterbankstrukturen.	41
3.2	In dieser Arbeit verwendete Simulationsparameter.	44
4.1	In dieser Arbeit verwendete Parametrierung der Schrittweitensteuerung. . .	55
4.2	In dieser Arbeit verwendete Parametrierung des Minimumschätzers. . . .	62
4.3	Parametrierung des Verfahrens zur Detektion von Raumänderungen mit zusätzlicher Schwelle.	64
4.4	Parametrierung des Verfahrens zur Detektion von Raumänderungen durch Vergleich der hochpass-/tiefpassgefilterten Fehlerleistung.	66
4.5	Parametrierung des Verfahrens zur Detektion von Raumänderungen mit- tels Vergleich von Mikrofon- und Fehlerleistung.	69
4.6	Parametrierung des Rettungsmechanismus.	70
4.7	Die erreichbaren maximalen stabilen Verstärkungen.	72
5.1	Die über den Sprachbereich und die Zeit gemittelten Leistungen nach der Maskierung.	93
6.1	Die Parameter der beiden getesteten Varianten.	107

Abkürzungen

AD/DA	Analog-Digital- bzw. Digital-Analog Wandler
DFT	Diskrete Fouriertransformation
EDC	<i>Energy Decay Curve</i> , Energieabklingkurve
HD	<i>Harmonic Distortion</i> , Klirrfaktor
ICC	<i>In-Car Communication</i> , Innenraumkommunikation
IDFT	Inverse diskrete Fouriertransformation
IIR	<i>Infinite Impulse Response</i> , Filter mit unendlicher Impulsantwort
LDS	Leistungsdichtespektrum
LMS	<i>Least-Mean-Square-Algorithmus</i>
MSV	Maximale stabile Verstärkung
NLMS	<i>Normalized-Least-Mean-Square-Algorithmus</i>
SNR	<i>Signal-to-Noise Ratio</i> , Signal-Rausch-Verhältnis

Notation

Konventionen und Operatoren

$(\cdot)^*$	Konjugiert komplexe Zahl
$ \cdot $	Betrag
$\angle(\cdot)$	Phase einer komplexen Zahl
$\hat{(\cdot)}$	Schätzwert
$E\{\cdot\}$	Erwartungswert
$*$	Faltungsoperation
\circ	Hadamardprodukt (elementweise Multiplikation)

$\overline{(\cdot)}$	Glättung
$\ \cdot\ ^2$	Euklidische Norm
$\lfloor \cdot \rfloor$	Runden auf die nächste ganze Zahl
$\lceil \cdot \rceil$	Aufrunden auf die nächste ganze Zahl
x	Zeitbereichssignale mit kleinen Buchstaben
X	Frequenzbereichssignale mit großen Buchstaben
\mathbf{x}, \mathbf{X}	Zeit- und Frequenzbereichsvektoren fett
\mathbf{x}^T	Vektor \mathbf{x} transponiert
\mathbf{x}^H	Vektor \mathbf{x} transponiert und konjugiert komplex
$\mathbf{x}(n) \circ \bullet \mathbf{X}(k)$	Diskrete Fouriertransformation

Wichtige Formelzeichen

n	Diskreter Zeitindex
k	Diskreter Blockindex
μ	Diskreter Frequenzindex
f	Kontinuierlicher Frequenzindex
e	Eulersche Zahl
j	Imaginäre Einheit
N	Länge der DFT
L	Rahmenversatz der Filterbank
M	Länge des adaptiven Filters im Teilband
m	Index der Filterkoeffizienten
$\mathbf{h}(n)$	Impulsantwort
N_h	Länge der Impulsantwort in Abtastwerten
$g(n), g(k)$	Verstärkung des ICC-Systems
$s(n), S(\mu, k)$	Lokale Sprache
$x(n), X(\mu, k)$	Lautsprechersignal
$y(n), Y(\mu, k)$	Mikrofonsignal
$e(n), E(\mu, k)$	Fehlersignal
$e_u(n), E_u(\mu, k)$	Ungestörtes Fehlersignal
$r(n), R(\mu, k)$	Rückkopplungsanteile am Mikrofon

$\hat{\mathbf{H}}(\mu, k)$	Geschätzte Impulsantwort im Teilband
$\alpha(\mu, k)$	Schrittweite des adaptiven Filters im Teilband
$\gamma(\mu, k)$	Geschätzter Systemabstand im Teilband
T_{60}	Nachhallzeit eines Raumes
T_D	Totzeit einer Impulsantwort

Kapitel 1

Einleitung

1.1 Innenraumkommunikationssysteme

Aufgrund hoher Hintergrundgeräuschpegel ist die Kommunikation in Personenkraftwagen oftmals beeinträchtigt. Das Hintergrundgeräusch setzt sich zusammen aus dem Motorgeräusch, den Abrollgeräuschen der Reifen und Windgeräuschen. Hinzu kommen die Geräusche der Lüftung im Inneren des Fahrzeugs. Mit zunehmender Geschwindigkeit steigt der Hintergrundgeräuschpegel im Inneren des Fahrzeugs an, was eine Anpassung der Sprache an die Geräuschumgebung erfordert. Dies ist für die Passagiere auf Dauer sehr anstrengend und ermüdend. Am stärksten von Hintergrundgeräuschen betroffen sind lange Fahrzeuge wie Vans oder größere Limousinen und offen fahrende Cabriolets. Selbst bei moderaten Geschwindigkeiten ist hier die Kommunikation zwischen den Insassen eingeschränkt.

Ein weiteres Problem neben den Hintergrundgeräuschen ist die für die Kommunikation ungünstige Sitzposition. So reden Fahrer und Beifahrer gegen die Windschutzscheibe, während die hinteren Passagiere gegen die Rückenlehne bzw. Kopfstütze der Vordersitze reden. Dies resultiert darin, dass sich die vorderen Insassen nach hinten drehen, um im Fond besser verstanden zu werden. Im Fall des Fahrers stellt dies ein großes Sicherheitsrisiko dar, da er sich dann nicht mehr auf die Straße vor ihm konzentrieren kann. Hinzu kommt der große Abstand zwischen den Sitzreihen in großen Fahrzeugen, wie langen Limousinen oder Vans mit mehr als zwei Sitzreihen. In solchen Fahrzeugen ist eine entspannte Kommunikation zwischen den vorderen Insassen und den Passagieren im Fond während der Fahrt nicht möglich. In [42] wird angegeben, dass der Schallpegel vom Mund eines Sprechers auf dem Fahrersitz bis zum Ohr eines Passagiers auf dem Rücksitz eine Dämpfung von 20 dB bis zu 30 dB erfährt.

Abhilfe schaffen sogenannte Innenraumkommunikationssysteme (engl. *In-Car Communication System*, ICC-System). In solchen Systemen wird die Sprache der Passagiere mit mundnahen Mikrofonen aufgenommen und anschließend verstärkt über Lautsprecher in der Nähe der Zuhörer wiedergegeben. Im einfachsten Fall werden hierfür die bereits im Fahrzeug verbauten Freisprechermikrofone sowie die Multimedialautsprecher verwendet. Es konnte in [38, 39] gezeigt werden, dass die Kommunikation im Fahrzeug mit einem solchen ICC-System deutlich verbessert wird. Als Maß für den Stress, dem die Passagiere aufgrund der Geräusche im Fahrzeug ausgesetzt sind, dienen in diesen Arbeiten linguistische Parameter. Insbesondere wird der Lombard-Effekt untersucht. Dieser besagt, dass ein Sprecher die charakteristischen Eigenschaften seiner Stimme an die Umgebungsgeräusche anpasst. Neben einer Erhöhung der Lautstärke gehört dazu z. B. auch

eine Erhöhung der Grundfrequenz bzw. der Tonhöhe bei zunehmendem Hintergrundgeräusch. Es konnte in einer umfangreichen Studie nachgewiesen werden, dass mit aktivem ICC-System der Lombard-Effekt bei den Sprechern deutlich weniger ausgeprägt ist, was auf eine entspanntere Kommunikation schließen lässt.

Aus den genannten Gründen werden ICC-Systeme mittlerweile von verschiedenen Automobilherstellern angeboten. An dieser Stelle sollen nur exemplarisch einige Systeme erwähnt werden. In den Fahrzeugen von Mercedes-Benz ist das System unter dem Namen *Sprachverstärkung* seit der V-Klasse BR447 und der S-Klasse X222 auf dem Markt verfügbar. Bei dem System handelt es sich um ein Einwegesystem, d. h. die Sprache wird nur von vorne nach hinten verstärkt. Volkswagen bietet ein ähnliches System ab dem Touran 2 und dem T5 an. Hier wird das System als *Digital Voice Enhancement* beworben. Ebenfalls ein Einwegesystem ist in den Fahrzeugen Highlander und Sienna von Toyota als *Driver Easy Speak* verbaut. Bei Land Rover und Jaguar wird das System *Conversation Assist* genannt und ist beispielsweise im Range Rover und im Jaguar XJ verfügbar. Die zunehmende Anzahl an verfügbaren Fahrzeugen mit ICC-Systemen zeigt erneut den Bedarf an solchen Systemen.

Beim Test der auf dem Markt verfügbaren Systeme stellt man allerdings fest, dass die Qualität oftmals unzureichend ist. Bei vielen der genannten Systeme ist die Verstärkung so gering, dass das System kaum wahrnehmbar ist und damit keinen Nutzen bringt. Bei anderen Systemen ist eine Sprachverstärkung zwar wahrnehmbar, jedoch klingt das verstärkte Sprachsignal aufgrund der verwendeten Signalverarbeitung sehr unnatürlich und wird selbst geringen Ansprüchen an die Klangqualität nicht gerecht. Der Hauptgrund für die geringe Verstärkung und die eingeschränkte Klangqualität ist die elektro-akustische Rückkopplung, die entsteht, wenn das Lautsprechersignal in das Mikrofon einkoppelt. Diese ist schematisch in Abb. 1.1 gezeigt. In der Abbildung ist ein ICC-System zu sehen, welches die Sprache im Fahrzeug von vorne nach hinten verstärkt. Man erkennt, dass sich am Mikrofon die Rückkopplung vom Lautsprecher und die lokale Sprache überlagern. Ist die Verstärkung des ICC-Systems zu groß, schwingt sich das System auf, was

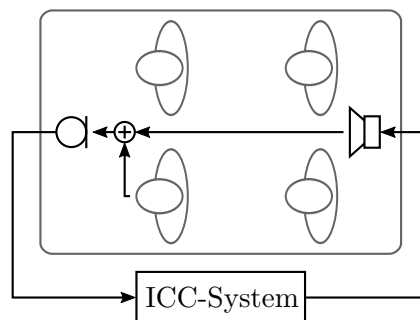


Abbildung 1.1: Block-Diagramm eines ICC-Systems. Am Mikrofon überlagern sich die Rückkopplung vom Lautsprecher und das lokale Sprachsignal.

als unangenehmes Pfeifen hörbar ist. Da dieses Pfeifen unbedingt vermieden werden soll, wird bislang zur Sicherheit die maximale Verstärkung begrenzt.

Das Problem der Instabilität rückgekoppelter elektro-akustischer Systeme ist technisch nach wie vor nicht umfassend gelöst und rechtfertigt daher auch die vorliegende Arbeit, in welcher deutliche Verbesserungen erreicht werden. Das Ziel dieser Arbeit ist es, die elektro-akustische Rückkopplung bei ICC-Systemen mittels digitaler Signalverarbeitung zu unterdrücken. Dabei soll der Fokus darauf liegen, im Vergleich zu bestehenden Systemen sowohl eine größere Verstärkung als auch eine bessere Klangqualität zu erzielen.

1.2 Inhalt dieser Arbeit

Im Anschluss an diese Einleitung werden in Kap. 2 die Randbedingungen im Fahrzeug beschrieben. Im Rahmen dieser Arbeit wurde ein Demonstratorfahrzeug aufgebaut. Bei diesem handelt es sich um den ebenfalls in Kap. 2 vorgestellten Mercedes-Benz Viano der Baureihe 639. Für dieses Fahrzeug wurde ein bidirektionales ICC-System entwickelt, welches die Kommunikation von vorne nach hinten und von hinten nach vorne unterstützt. Die Randbedingungen werden in Form von Messungen vorgestellt, welche in dem Fahrzeug durchgeführt wurden. Die Messungen umfassen insbesondere die Impulsantworten der relevanten akustischen Pfade sowie der Hintergrundgeräusche. Die vorgestellten Messergebnisse dienen im weiteren Verlauf der Arbeit als Grundlage für die Entwicklung der Algorithmen. Um die Algorithmen im Fahrzeug im realen Betrieb testen zu können, wurde das Software-Rahmenwerk RACE entwickelt, in welches die Echtzeitimplementierungen der Algorithmen integriert sind. Das Rahmenwerk wird in Anhang B vorgestellt.

In Kap. 3 werden die für die Arbeit benötigten mathematischen Grundlagen hergeleitet. Zunächst wird das Stabilitätskriterium für die geschlossene elektro-akustische Schleife erläutert. Nach einem Überblick über den Stand der Technik werden zwei grundsätzliche Lösungsansätze zur Verbesserung der Stabilität vorgestellt. Dies ist zum einen die Kompensation der Rückkopplung mittels adaptiver Filter und zum anderen die Reduktion der Rückkopplung mit einem Wiener-Filter. Auf diesen beiden Ansätzen wird im weiteren Verlauf der Arbeit aufgebaut. Am Ende des Kapitels wird die verwendete Filterbankstruktur erläutert.

Kap. 4 stellt den Kern dieser Arbeit dar. In dem Kapitel wird eine neuartige Schrittweitensteuerung für die adaptive Rückkopplungskompensation vorgestellt. Mit dem vorgestellten Verfahren können im ICC-System des Demonstratorfahrzeugs Verstärkungen erzielt werden, die sehr weit über denen aktueller Systeme liegen, bei gleichzeitig verbesserter Klangqualität. Die Inhalte dieses Kapitels basieren auf diversen Veröffentlichungen, die im Rahmen dieser Arbeit entstanden sind. Die grundlegende Idee der Schrittweitensteuerung wird in [6] beschrieben. In [7] wird das Verfahren um eine Verstärkungsregelung erweitert. Das für diese Verstärkungsregelung benötigte Verfahren zur Schätzung des Systemabstands wird in [8] beschrieben.

Die Rückkopplungskompensation wird in Kap. 5 um ein Postfilter erweitert. Die Aufgabe dieses Postfilters ist es, diejenigen Rückkopplungsanteile, welche nach der Rückkopplungskompensation übrig bleiben, mit einem Wiener-Filter zu reduzieren. Ziel ist es dabei, einen trockenen Klang zu erreichen, welcher keine rückkopplungsbedingten

Nachhallartefakte mehr aufweist. Das beschriebene Verfahren wurde in [24] publiziert.

Kap. 6 behandelt zwei Verfahren zur Verbesserung der Sprachqualität und der Sprachverständlichkeit. Das erste Verfahren ist ein Algorithmus zur automatischen Entzerrung des Lautsprechersignals am Zuhörerohr. Hierdurch soll der Eindruck eines linearen Frequenzgangs entstehen, was in einem natürlichen Klangerlebnis resultiert. Die entsprechende Veröffentlichung ist in [9] zu finden. Das zweite Verfahren verwendet nichtlineare Kennlinien, um ein Signal mit Harmonischen anzureichern. Auf diese Weise kann die Sprachverständlichkeit in einer stark geräuschbehafteten Umgebung verbessert werden. Der Algorithmus wurde in [10] publiziert.

Kap. 7 fasst schließlich die Ergebnisse der Arbeit zusammen und gibt einen Ausblick auf weiterführende Themen.

Zusätzlich zu den oben genannten Inhalten sind im Rahmen dieser Arbeit zwei weitere Veröffentlichungen entstanden, welche allerdings nicht in die vorliegende Arbeit eingeflossen sind. Zum einen handelt es sich hierbei um den in [41] vorgestellten De-Esser. Mit diesem werden störende Zischlaute im Sprachsignal reduziert. Die Frequenzen dieser Zischlaute sind sprecherabhängig und können durch den Frequenzgang des Systems unnatürlich hervorgehoben werden. Das Verfahren detektiert diese Zischlaute unabhängig vom Eingangspegel und dämpft diese mit zwei verschiedenen Kerbfiltern. Zum anderen wurde ein Mercedes-Benz S-Klasse Cabriolet (A217) mit Gurtmikrofonen und Kopfstützenlautsprechern ausgestattet und ein ICC-System integriert. In diesem Fahrzeug unterstützt das bidirektionale ICC-System die Kommunikation zwischen Fahrer und Beifahrer. Dieses System wird in [5] detailliert beschrieben.

Kapitel 2

Randbedingungen im Fahrzeug

Sämtliche in dieser Arbeit betrachteten Aspekte werden im Hinblick auf den Einsatzzweck in Personenkraftwagen untersucht. Daher wird in diesem Kapitel zunächst auf die besonderen Gegebenheiten und Anforderungen im Fahrzeug eingegangen. Im ersten Teil des Kapitels wird das Demonstratorfahrzeug beschrieben, welches für die praktischen Tests aufgebaut worden ist. Anschließend werden die Randbedingungen in diesem Fahrzeug beschrieben. Hierzu gehören insbesondere die akustischen Eigenschaften des Fahrzeuginnenraums, welche in dieser Arbeit durch die Impulsantworten von den Lautsprechern zu den jeweiligen Mikrofonen beschrieben werden, sowie die spektralen Eigenschaften der Hintergrundgeräusche. Auf Basis dieser im Fahrzeug gegebenen Randbedingungen werden die Anforderungen an die Signalverarbeitung festgelegt.

2.1 Demonstratorfahrzeug

Im Rahmen dieser Arbeit wurde ein Demonstratorfahrzeug aufgebaut, mit dem es ermöglicht wurde die entwickelten Algorithmen unter realen Bedingungen testen. Dafür stand ein Mercedes-Benz Viano zur Verfügung (vgl. Abb. 2.1). Dieses Fahrzeug ist mit drei Sitzreihen ausgestattet und aufgrund seiner Größe gut geeignet um den Mehrwert eines ICC-Systems aufzuzeigen. Das Fahrzeug ist serienmäßig lediglich mit einem Mikrofon in der Dachbedieneinheit sowie Lautsprechern in den vorderen Türen und der hinteren Seitenverkleidung ausgerüstet. Um verschiedenste Konfigurationen testen zu können, wurden zusätzliche Mikrofone und Lautsprecher im Dachhimmel verbaut, sodass das Demonstratorfahrzeug mit insgesamt acht Lautsprechern und vier Mikrofonen ausgestattet ist. Zwei Lautsprecher befinden sich in den Türen vorne links und rechts, zwei weitere in der Seitenverkleidung auf Kniehöhe der dritten Sitzreihe. Die vier zusätzlichen Lautsprecher sind im Dachhimmel direkt über den Sitzplätzen von Fahrer und Beifahrer sowie den Passagieren der dritten Sitzreihe verbaut. Da sich diese Lautsprecher sehr nahe an den Ohren der Insassen befinden, wird ihr Pegel im Vergleich zu den seitlichen- bzw. den Tür-Lautsprechern um -3 dB abgesenkt. Die vier Mikrofone sind 32 cm vor den Dachlautsprechern, ebenfalls im Dachhimmel, positioniert. In Abb. 2.2 sind die genauen Lautsprecher- und Mikrofonpositionen im Fahrzeug dargestellt.

Abb. 2.3 zeigt ein Bild der Sitzplätze von Fahrer und dem Passagier hinten links. Gut zu erkennen sind die Lautsprecher und Mikrofone im Dachhimmel sowie der Lautsprecher in der hinteren Seitenverkleidung. Die Tür-Lautsprecher vorne werden auf dem Bild vom Lenkrad verdeckt.



Abbildung 2.1: Als Demonstratorfahrzeug wurde ein Mercedes-Benz Viano der Baureihe 639 mit zusätzlichen Lautsprechern und Mikrofonen ausgerüstet.

Die verbauten Lautsprecher und Mikrofone sind über Verstärker bzw. Mikrofon-Vorverstärker mit einer Soundkarte verbunden. Diese ist wiederum an einen PC angeschlossen, auf dem die Signalverarbeitung berechnet wird. Hierfür wurde das Audio-Rahmenwerk RACE entwickelt, in welches die Signalverarbeitungsalgorithmen integriert sind. Das Rahmenwerk wird in Anhang B vorgestellt.

2.2 Messung der Impulsantworten im Fahrzeug

Verantwortlich für die elektro-akustische Rückkopplung sind die Pfade zwischen den Lautsprechern und den Mikrofonen. Diese Pfade können durch Impulsantworten beschrieben werden. Aus der Impulsantwort können Informationen über die Akustik des untersuchten Raumes abgeleitet werden. Wird die Impulsantwort in den Frequenzbereich transformiert, erhält man den Frequenzgang. Dieser beschreibt die frequenzabhängige Dämpfung, die ein Signal auf dem Weg von den Lautsprechern zum Mikrofon erfährt. Ebenfalls abgeleitet werden können die Totzeit und die Nachhallzeit. Die Totzeit T_D beschreibt die Laufzeit, die der Schall benötigt, um auf dem direkten Weg vom Lautsprecher zum Mikrofon zu gelangen. Zeitlich nach dem Direktschall treffen die Anteile am Mikrofon ein, welche Reflexionen im Raum erfahren haben. Da diese Nachhall-Anteile einen längeren Weg zurückgelegt haben, sind sie schwächer als der Direktschall. Der Zeitpunkt, ab dem der Nachhall im Vergleich zum Direktschall auf -60 dB abgeklungen ist, wird als Nachhallzeit T_{60} bezeichnet. Nachhallzeit und Totzeit können aus der Energieabklingkurve (engl. *Energy Decay Curve*, EDC) der Impulsantwort berechnet werden. Die EDC

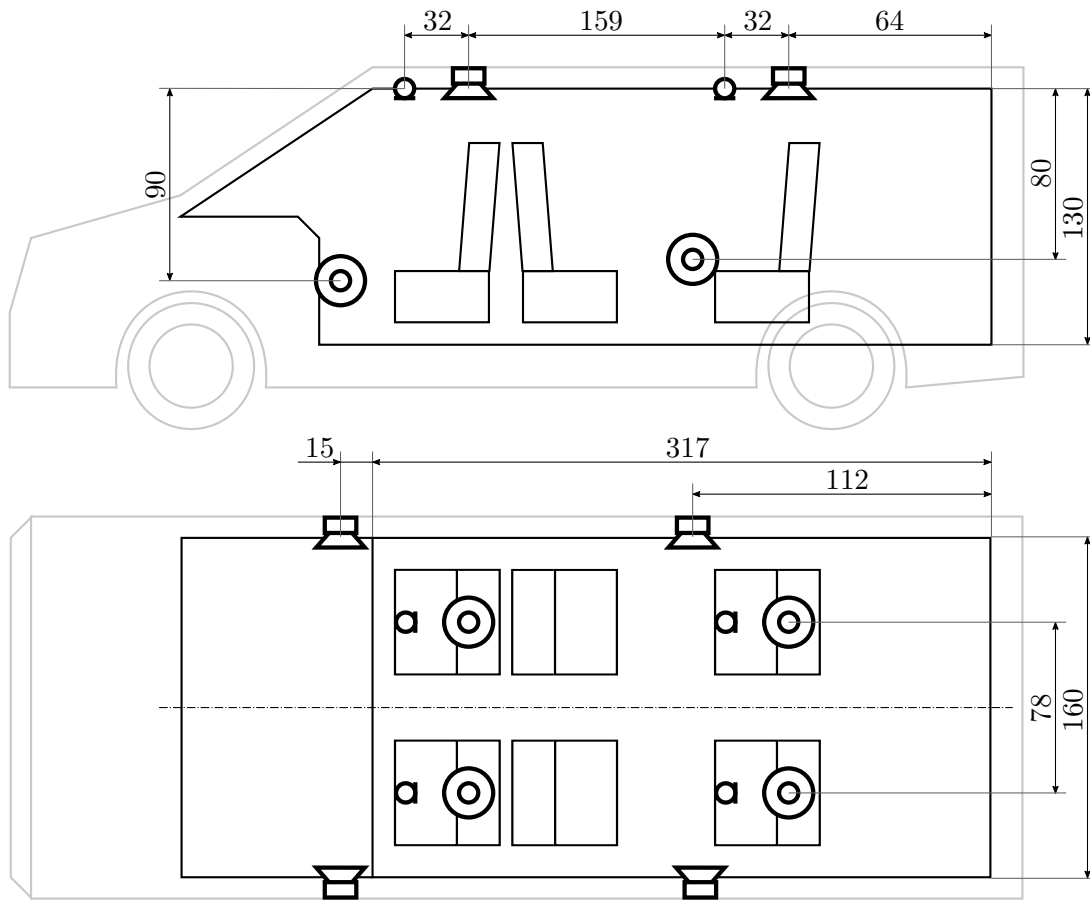


Abbildung 2.2: Lautsprecher-/Mikrofonpositionen im Demonstratorfahrzeug (alle Maße in cm, Zeichnung nicht maßstäblich).

berechnet sich aus der Rückwärtsintegration des Betragsquadrates der Impulsantwort [57]

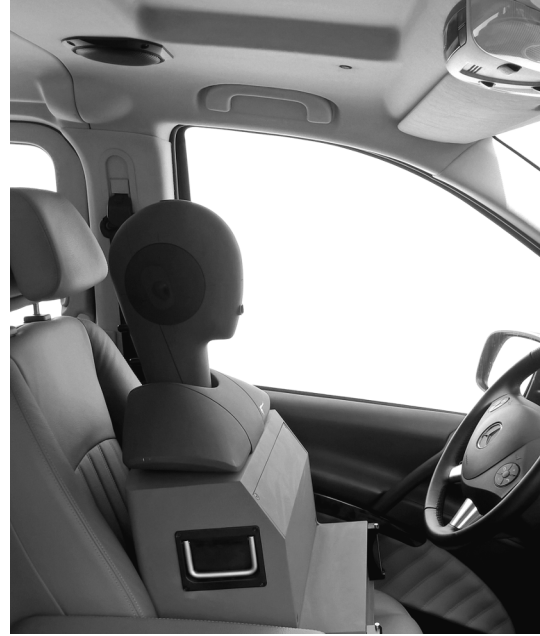
$$\text{EDC}(n) = \frac{\sum_{i=n}^{N_h-1} |h_i|^2}{\sum_{i=0}^{N_h-1} |h_i|^2}. \quad (2.1)$$

Hierbei entspricht n dem Index der diskreten Zeit und h_i den Abtastwerten der gemessenen, zeitdiskreten Impulsantwort. N_h ist die Länge der Impulsantwort in Abtastwerten. Zur Bestimmung der Totzeit wird der Zeitpunkt verwendet, ab dem die EDC zu fallen beginnt. Von dem Zeitpunkt, ab dem die EDC unter -60 dB gefallen ist, wird T_D abgezogen, um T_{60} zu erhalten.

Nachfolgend wird kurz die Messung der Impulsantworten beschrieben. Die auf diese Weise im Fahrzeug gemessenen Impulsantworten werden im weiteren Verlauf dieser



(a) Dritte Sitzreihe.



(b) Fahrersitz.

Abbildung 2.3: Ein Kunstkopf im Demonstratorfahrzeug. Zu sehen sind die Mikrofone und Lautsprecher im Dachhimmel sowie in der Seitenverkleidung hinten. Die Lautsprecher in der Fahrertür werden vom Lenkrad verdeckt.

Arbeit für die Simulationen verwendet.

2.2.1 Methode

Als Signal für die Impulsantwortmessung dient ein in seiner Frequenz linear abfallendes Sinussignal, dessen Verlauf empirisch bestimmt worden ist. Die Abtastrate beträgt $f_s = 44.100$ Hz, die Frequenz des Sinussignals fällt innerhalb 2^{13} Abtastwerten (entspricht ca. 186 ms) von $f_s/2$ auf 0 Hz. Der Frequenzabfall ist linear, jedoch mit einer geringeren negativen Steigung für die tiefen Frequenzen unterhalb 1100 Hz. Der Grund für den langsameren Abfall ist die längere Einschwingzeit des Systems bei tiefen Frequenzen. Das Spektrogramm dieses Sinussweeps ist in Abb. 2.4 dargestellt.

Für die Messung wird dieses Signal 100 mal nacheinander über die Lautsprecher im Fahrzeug ausgegeben und mit den Fahrzeugmikrofonen wieder aufgenommen. Da die Soundkarte aufgrund der Blockverarbeitung und den AD-/DA Wandlern eine Laufzeit einfügt, wird zusätzlich ein Referenzsignal aufgenommen. Dazu wird das Signal vom Soundkartenausgang direkt wieder auf einen Soundkarteneingang geleitet. Das Prinzip der Messung ist schematisch in Abb. 2.5 dargestellt. Aus dem Referenzsignal und dem vom jeweiligen Fahrzeugmikrofon aufgezeichneten Messsignal kann die Impulsantwort zwischen Lautsprecher und Mikrofon bestimmt werden. Wichtig für das Verständnis ist, dass die Impulsantwort in diesem Fall nicht nur das Übertragungsverhalten des

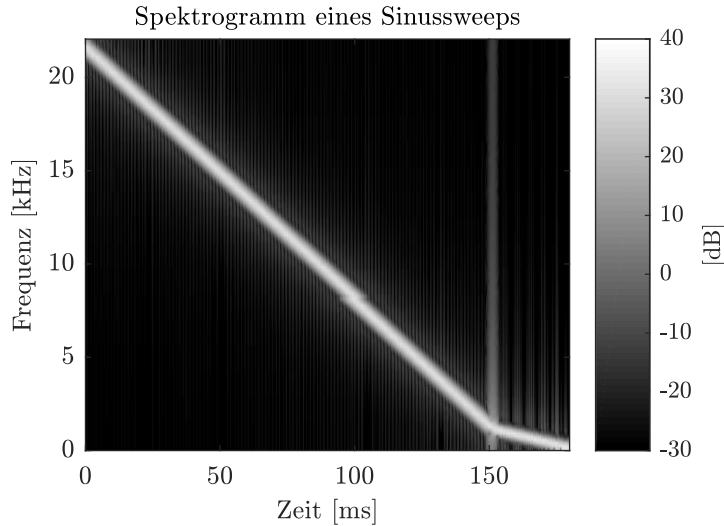


Abbildung 2.4: Das Spektrogramm eines Sinussweeps. Für die Impulsantwortmessung werden 100 dieser Sweeps nacheinander abgespielt.

Raumes beinhaltet. Dadurch, dass die Impulsantwort mit den im Fahrzeug verbauten Lautsprechern und Mikrofonen gemessen wird, wird auch das Übertragungsverhalten dieser elektrischen Komponenten mitgemessen.

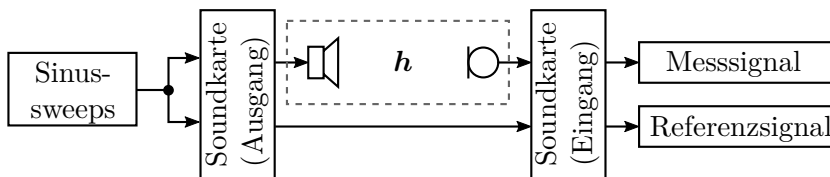


Abbildung 2.5: Schematischer Ablauf der Impulsantwortmessung. Gemessen wird die Impulsantwort h von Lautsprecher zu Mikrofon. Es gilt zu beachten, dass h auch das Übertragungsverhalten von Lautsprecher und Mikrofon selbst beinhaltet.

Zur Berechnung der Impulsantwort wird das mit dem Mikrofon aufgenommene Messsignal zunächst mit einem nullphasigen FIR Bandpass Filter der Ordnung 2048 gefiltert. Der Frequenzgang des Filters hat im Bereich 80 Hz bis 20 kHz den Betrag 0 dB. Außerhalb des Durchlassbereichs dämpft das Filter mit -20 dB. Um den Einfluss von Einschwingvorgängen zu minimieren, werden vom Messsignal und vom Referenzsignal die ersten beiden Sinussweeps abgeschnitten. Danach wird das Kreuzleistungsdichtespektrum zwischen Messsignal und Referenzsignal $P_{xy}(f)$ sowie das Autoleistungsdichtespektrum des Referenzsignals $P_{yy}(f)$ bestimmt. Der Quotient von Kreuzleistungsdichtespektrum und

Autoleistungsdichtespektrum ist die Übertragungsfunktion $H(f)$

$$H(f) = \frac{P_{xy}(f)}{P_{yy}(f)}, \quad (2.2)$$

wobei f die kontinuierliche Frequenz bezeichnet. Wird $H(f)$ in den Zeitbereich transformiert, erhält man schließlich die Impulsantwort [36].

2.2.2 Ergebnisse

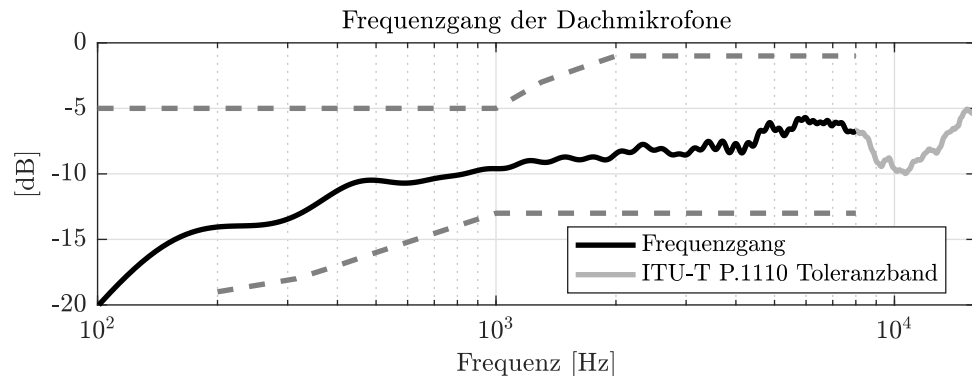


Abbildung 2.6: Der Frequenzgang der verwendeten Mikrofone gemessen im Freifeld und das Toleranzband aus der ITU-T P.1110 Empfehlung.

Wie im vorangegangenen Abschnitt beschrieben, beinhalten die im Fahrzeug gemessenen Impulsantworten auch das Übertragungsverhalten der Lautsprecher und Mikrofone. Die in dieser Arbeit verwendeten Mikrofone sind für Breitband-Freisprechsysteme in Kraftfahrzeugen spezifiziert. Die entsprechende Empfehlung ist die ITU-T P.1110 [34]. In dieser wird ein Toleranzband im Bereich von 100 Hz bis 8000 Hz vorgegeben. Eine Empfehlung der ITU-T für ICC-Systeme ist unter der Bezeichnung P.ICC in Arbeit, wurde allerdings bislang noch nicht veröffentlicht. Abb. 2.6 zeigt den im Freifeld gemessenen Frequenzgang der Mikrofone. Es ist zu erkennen, dass dieser innerhalb des ITU-T Toleranzbandes liegt. Unterhalb 200 Hz fällt der Frequenzgang steil ab. Dieser Abfall ist beabsichtigt, da sich die Energie des Fahrgeräusches vor allem bei tiefen Frequenzen konzentriert. Daher werden die Mikrofone so ausgelegt, dass sie diesen Frequenzbereich dämpfen. Oberhalb 8000 Hz ist der Frequenzgang nicht spezifiziert, weshalb dieser Bereich bis 16 kHz grau dargestellt ist. Bei 10 kHz ist eine Absenkung um ca. 4 dB zu sehen. Aufgrund des Mikrofonfrequenzgangs sind im weiteren Verlauf der Arbeit in den Bildern nur Frequenzen bis 16 kHz dargestellt.

In Abb. 2.7 sind die Frequenzgänge von der Summe der hinteren- bzw. vorderen Lautsprecher zum jeweiligen vorderen- bzw. hinteren Mikrofon gezeigt. Generell ist zu sehen, dass die Frequenzgänge tieffrequent stark abfallen. Dies liegt wie erläutert am mitgemessenen Frequenzgang der verwendeten Mikrofone. Man erkennt außerdem, dass die

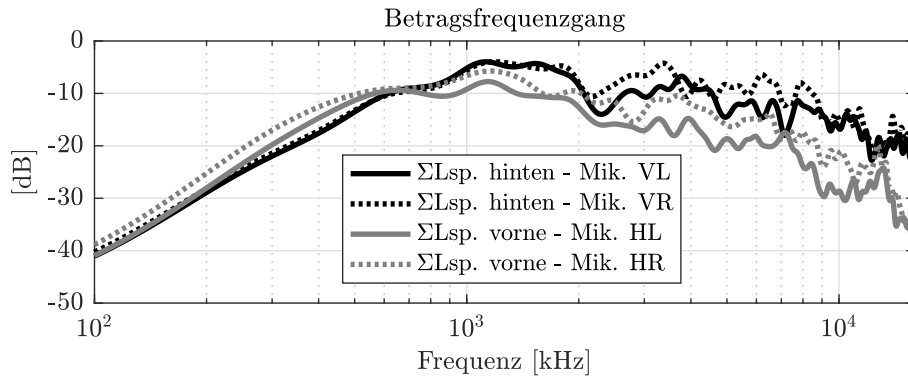


Abbildung 2.7: Die Frequenzgänge von den hinteren Lautsprechern auf die vorderen Mikrofone und von den vorderen Lautsprechern auf die hinteren Mikrofone. Das Fahrzeug ist leer.

Frequenzgänge von den hinteren Lautsprechern zum Fahrermikrofon (schwarze, durchgezogene Linie) bzw. Beifahrermikrofon (schwarze, gestrichelte Linie) weitgehend übereinander liegen. Dies liegt daran, dass der Fahrgastraum zur Längsachse symmetrisch ist und daher auch die Pfade näherungsweise gleich sind. Im Vergleich dazu sind die Frequenzgänge von den vorderen Lautsprechern zu den hinteren Mikrofonen vor allem hochfrequent gedämpft. Der Grund hierfür ist, dass die Lautsprecher in den Türen relativ weit unten positioniert sind und durch die vorderen Sitze nach hinten abgeschirmt sind. Es ist außerdem zu erkennen, dass der Frequenzgang zum linken hinteren Mikrofon (graue, durchgezogene Linie) stärker gedämpft ist, als zum rechten Mikrofon (graue, gestrichelte Linie). Auch hier ist der Grund die Einbauposition der Tür Lautsprecher. Der Schall des linken Lautsprechers wird durch das Lenkrad und das vorstehende Armaturenbrett stärker gedämpft als auf der Beifahrerseite.

In Abb. 2.8 sind die EDCs für die beschriebenen Pfade dargestellt. Im Fahrzeug fällt der Schalldruck exponentiell gegen null ab, d. h. auch die EDC läuft theoretisch gegen null. Es ist allerdings zu sehen, dass in der Abbildung die EDC ab ca. -40 dB abgeknickt ist und nur noch langsam fällt. Ab diesem Punkt überwiegt das Messrauschen der Mikrofone, welches verhindert, dass die berechnete EDC auf null abfällt. Aus diesem Grund wird zur Ermittlung der Nachhallzeit eine Regressionsgerade durch den Bereich der EDC gelegt, in dem die Kurve von -10 dB auf -30 dB abfällt (in Abb. 2.8 fett, schwarz hervorgehoben). Aufgrund der Regressionsgeraden kann die bestimmte Nachhallzeit nur als grober Richtwert verstanden werden. Für die untersuchten Pfade liegt sie etwa zwischen 116,6 ms und 120,2 ms.

Die Totzeit von den hinteren Lautsprechern zu den vorderen Mikrofonen wurde zu $T_D = 5,7$ ms ermittelt. Bei einer Schallgeschwindigkeit von $c = 343$ m/s entspricht dies einer Distanz von

$$5,7 \cdot 10^{-3} \cdot 343 \text{ m} \approx 1,96 \text{ m.} \quad (2.3)$$

Der kürzeste Weg, den der Schall von den hinteren Lautsprechern zum linken bzw. rechten vorderen Mikrofon zurücklegt, ist der zwischen hinterem seitlichem Lautsprecher

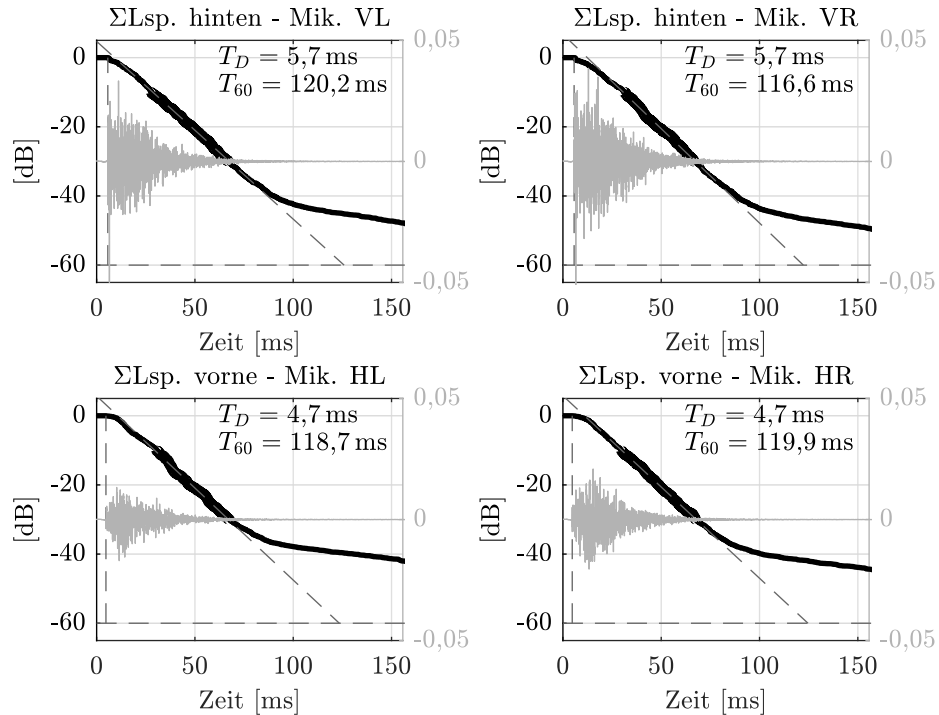


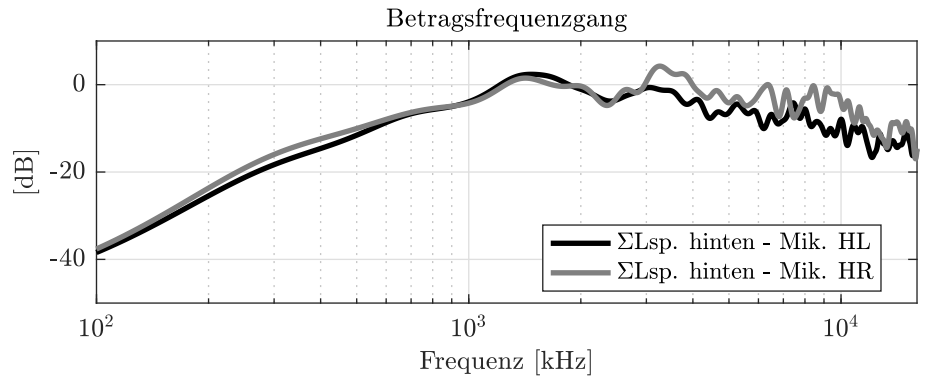
Abbildung 2.8: Energieabklingkurven der gemessenen Impulsantworten zur Bestimmung von Nachhallzeit T_{60} und Totzeit T_D .

und dem jeweiligen Mikrofon auf der gleichen Seite. Aus der Geometrie in Abb. 2.2 geht hervor, dass diese Distanz etwa 1,97 m beträgt, was in guter Näherung der aus der Totzeit berechneten Distanz entspricht. Analog kann die bei der Übertragung von vorne nach hinten ermittelte Totzeit (4,7 ms) verifiziert werden. Der kürzeste Abstand ist hier zwischen vorderem Dachlautsprecher und hinterem Mikrofon (1,59 m). Der aus der Totzeit berechnete Abstand beträgt

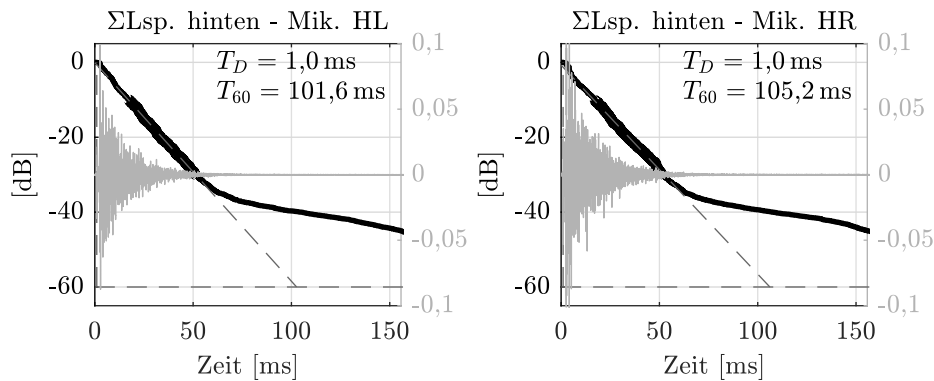
$$4,7 \cdot 10^{-3} \cdot 343 \text{ m} \approx 1,61 \text{ m.} \quad (2.4)$$

Auch hier stimmen die Werte gut überein.

In einem bidirektionalen System, welches die Sprache sowohl von vorne nach hinten als auch von hinten nach vorne verstärkt, sind die vorderen und die hinteren Mikrofone aktiv. Dies hat zur Folge, dass die hinteren Lautsprecher auch in die hinteren Mikrofone einkoppeln und äquivalent dazu die vorderen Mikrofone in die vorderen Lautsprecher. Exemplarisch hierfür ist in den Abb. 2.9a bzw. 2.9b die hintere Mikrofon/Lautsprecher-Kopplung gezeigt. Die Betragsfrequenzgänge liegen hierbei weitgehend übereinander. Es ist allerdings im Vergleich zu den in Abb. 2.7 gezeigten langen Pfaden zu erkennen, dass die Beträge hier insgesamt größer sind. Dies liegt an der kürzeren Distanz zwischen den Lautsprechern und den Mikrofonen, was dazu führt, dass die Lautsprecher stärker in die Mikrofone einkoppeln. Da die Dachlautsprecher nach Abb. 2.2 nur 32 cm von den jeweiligen Dachmikrofonen entfernt sind, ist hier auch die Totzeit T_D deutlich geringer als



(a) Betragsfrequenzgänge.



(b) Energieabklingkurven.

Abbildung 2.9: Die Frequenzgänge und zugehörigen Impulsantworten, bzw. Energieabklingkurven von der Summe der hinteren Lautsprecher zu den beiden hinteren Mikrofonen.

bei den Impulsantworten aus Abb. 2.8. Zur Überprüfung der Plausibilität des gemessenen Wertes wird erneut aus der Totzeit auf den Abstand zurückgerechnet, was hier auf

$$1,0 \cdot 10^{-3} \cdot 343 \text{ m} \approx 0,34 \text{ m} \quad (2.5)$$

führt. Auch dieser Wert weicht nur 2 cm vom gemessenen Abstand ab. Die Unterschiede in der Nachhallzeit T_{60} zwischen linkem und rechtem Mikrofon sind ebenfalls auf Messungenauigkeiten zurückzuführen.

Die bisher beschriebenen akustischen Eigenschaften beschreiben das leere Fahrzeug ohne Passagiere. Wird das Fahrzeug beladen, ändert sich die akustische Umgebung, was sich vor allem auf den Frequenzgang und die Nachhallzeit auswirkt. Sehr deutlich ist dies zu sehen, wenn sich sechs Passagiere im Fahrzeug befinden, d. h. jeder Sitzplatz besetzt ist. Abb. 2.10a stellt den Frequenzgang des leeren Fahrzeuges dem des voll besetzten Fahrzeuges gegenüber. Die beiden Frequenzgänge beschreiben jeweils den Pfad von der Summe der hinteren Lautsprecher zum Fahrermikrofon. Es ist zu sehen, dass beim voll besetzten Fahrzeug das Signal über den gesamten Frequenzbereich stärker ge-

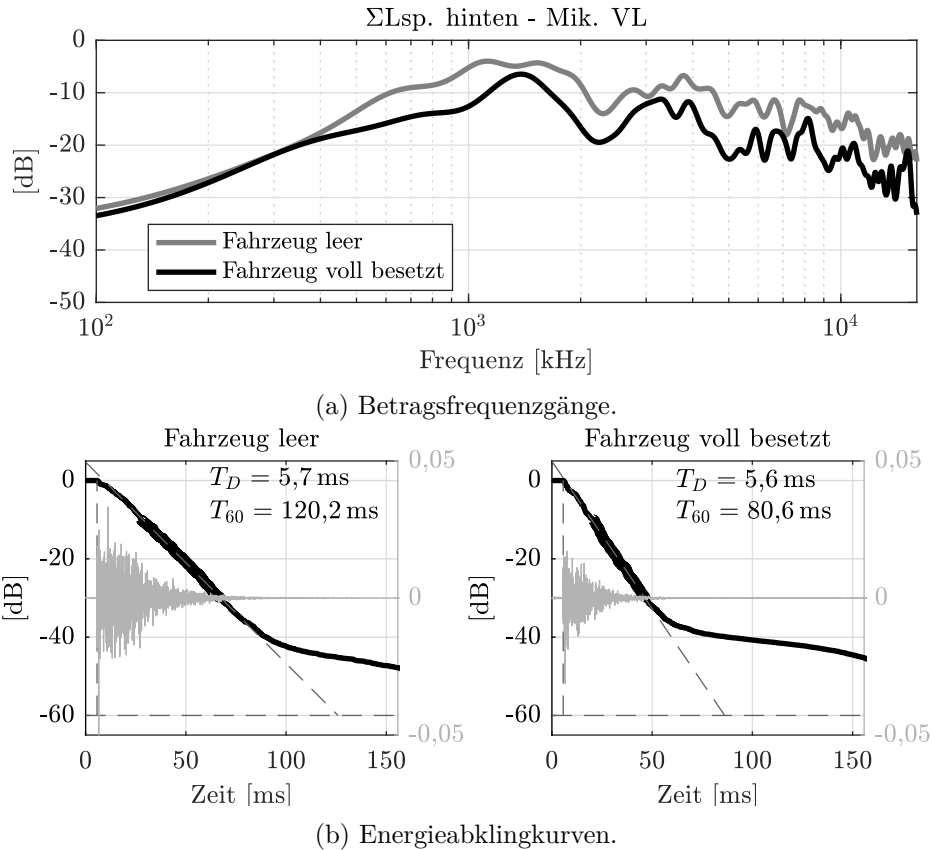


Abbildung 2.10: Vergleich der Impulsantworten von den hinteren Lautsprechern zum Fahrermikrofon bei leerem und voll besetztem Fahrzeug. Die unterschiedliche Totzeit T_D ist auf Messungenauigkeiten zurückzuführen.

dämpft wird als beim leeren Fahrzeug. Insbesondere hochfrequente Schallanteile werden von den weichen Kleidungsstücken der Passagiere absorbiert. Deutlich wird dies auch bei der Nachhallzeit. Aufgrund der stärkeren Absorption werden weniger Schallanteile reflektiert. In Abb. 2.10b ist dementsprechend die Nachhallzeit beim voll besetzten Fahrzeug um etwa 40 ms kürzer als beim leeren Fahrzeug.

Die akustischen Eigenschaften werden jedoch nicht nur durch die Beladung geändert. Auch Bewegungen der Passagiere können Änderungen hervorrufen. Bauartbedingt sind z. B. die hinteren seitlichen Lautsprecher auf Höhe der Knie der Passagiere in der dritten Sitzreihe positioniert. In Abb. 2.11 sind die Frequenzgänge gezeigt, wenn der Passagier normal sitzt und wenn er sein Knie vor den Lautsprecher hält. Deutlich zu sehen ist, dass mit vorgehaltenem Knie einzelne Spitzen (z. B. bei ca. 1,5 kHz, 3,4 kHz und 7,8 kHz) um etwa -4 dB gedämpft werden. An der Nachhallzeit und der Totzeit ändert sich in diesem Szenario nichts. Noch deutlicher ändert sich der Frequenzgang, wenn der Fahrer seine Hand vor das Fahrermikrofon hält und diese wieder wegnimmt. Dies ist in Abb. 2.12 gezeigt. Hier erkennt man, dass durch die vorgehaltene Hand bei 1,2 kHz eine Resonanz

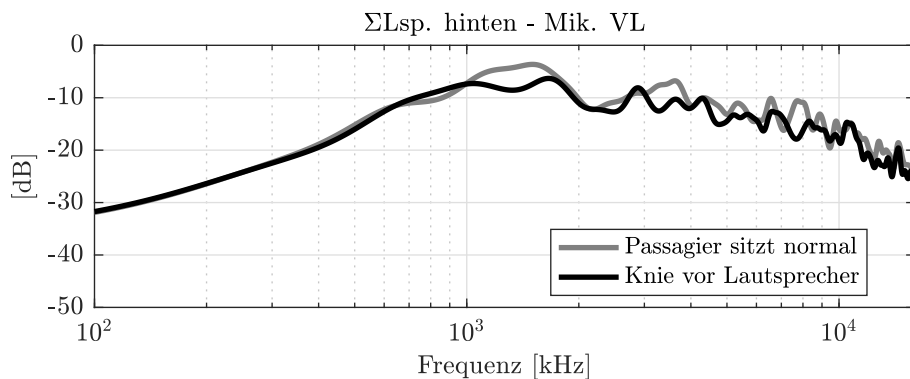


Abbildung 2.11: Der Frequenzgang, wenn der Passagier auf dem linken Sitz der dritten Sitzreihe sein Knie direkt vor den linken seitlichen Lautsprecher hält. Zum Vergleich der Frequenzgang, wenn der Passagier normal sitzt.

entsteht. Bei dieser Frequenz hat die schwarze Kurve eine deutliche Überhöhung. Danach fällt der Frequenzgang stark ab und ist bei den höheren Frequenzen um ca. 10 dB gedämpft im Vergleich zu dem Fall ohne Hand.

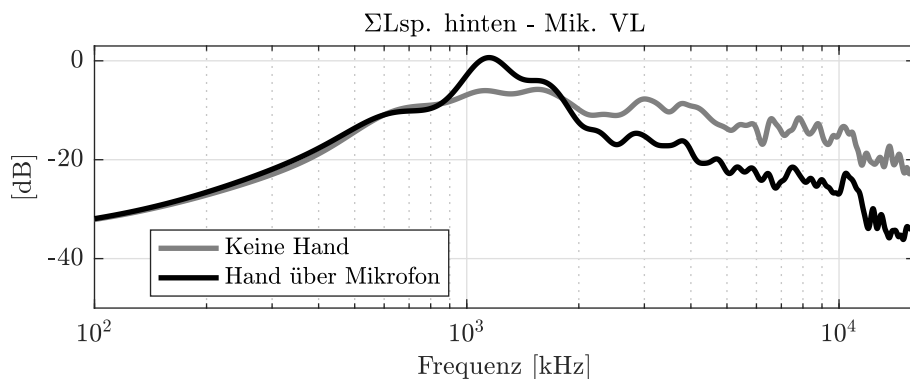


Abbildung 2.12: Der Frequenzgang, wenn der Fahrer mit seiner Hand das Mikrofon verdeckt und diese wieder wegnimmt.

2.3 Anforderungen an die Signalverarbeitung

In dem im vorangegangenen Abschnitt beschriebenen Demonstratorfahrzeug wurde das in Abb. 2.13 dargestellte bidirektionale ICC-System aufgebaut. Dieses unterstützt die Sprachkommunikation von vorne nach hinten sowie von hinten nach vorne. Dabei wird eine Pegelwaage eingesetzt, um zwischen den beiden Wegen umzuschalten, sodass zu jedem Zeitpunkt immer nur ein Weg aktiv ist (vgl. z. B. [55]). Dies rechtfertigt, dass in den nachfolgenden Kapiteln jeweils nur ein Weg isoliert betrachtet wird. Die vier vorderen bzw. hinteren Lautsprecher sind dabei zu einem Kanal zusammengefasst, die

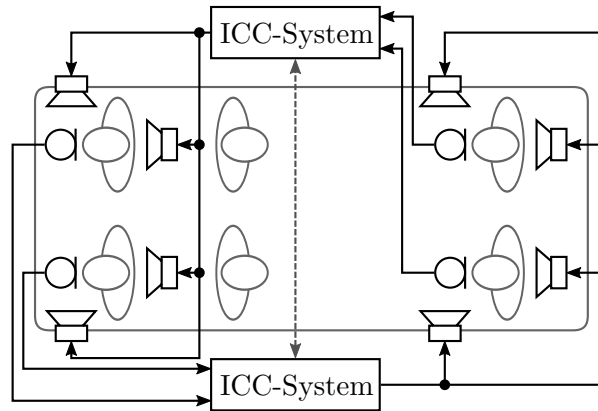


Abbildung 2.13: Block-Diagramm des bidirektionalen ICC-Systems welches im beschriebenen Demonstratorfahrzeug aufgebaut wurde.

Rückkopplung wird an einem der beiden den Lautsprechern gegenüberliegenden Mikrofonen betrachtet. Damit reduziert sich das System auf das bereits in Abb. 1.1 auf S. 2 gezeigte Blockdiagramm.

Die vier wichtigsten Anforderungen, welche an die Signalverarbeitung eines solchen ICC-Systems gestellt werden, lassen sich wie folgt zusammenfassen:

- Die Signalverarbeitung muss die Rückkopplung unterdrücken, damit das System auch bei hohen Verstärkungen stabil bleibt.
- Die Signalverarbeitung muss sich selbstständig an veränderte Gegebenheiten wie Raumänderungen oder verschiedene Sprecher anpassen, damit ein gleichbleibendes Klangerlebnis gewährleistet werden kann.
- Die Signalverarbeitung muss eine geringe Laufzeit haben, da sonst die Verzögerung zwischen Direktschall und Lautsprechersignal hörbar wird.
- Die Signalverarbeitung muss Hintergrundgeräusche unterdrücken und bei zunehmendem Geräuschpegel die Verstärkung des Systems automatisch anpassen.

Speziell auf die Problematik der elektro-akustischen Rückkopplung wird in Kap. 3 im Detail eingegangen. Die verbleibenden drei Anforderungen werden im Folgenden kurz erläutert.

2.3.1 Automatische Anpassung

Eine Änderung des akustischen Raumes beeinflusst direkt auch das klangliche Ergebnis am Zuhörerohr. Solche Änderungen können, wie in Abschn. 2.2 beschrieben, durch Bewegungen der Passagiere zustande kommen. Darüber hinaus spielen jedoch auch Toleranzen bei der Fertigung eine große Rolle. Insbesondere bei der Fertigung und der Montage der Mikrofone und der Lautsprecher haben kleine Abweichungen zum Teil große Auswirkungen auf die Klangeigenschaften. Zudem nehmen auch Alterungsprozesse der Bauteile

sowie Temperaturunterschiede Einfluss auf den Klang. Zusätzlich zu den akustischen Eigenschaften des Raumes variieren Stimmlage und Lautstärke bei verschiedenen Sprechern. All dies führt dazu, dass sich die Signalverarbeitungsalgorithmen adaptiv zur Laufzeit an veränderte Gegebenheiten oder unterschiedliche Sprecher anpassen müssen.

2.3.2 Geringe Laufzeit

Kritisch ist auch die Anforderung an die Laufzeit des Systems. Der Zuhörer nimmt sowohl den Direktschall als auch das verarbeitete Lautsprechersignal wahr. Je größer die Verzögerung zwischen Direktschall und Lautsprechersignal ist, desto halliger ist der Klangeindruck des Systems beim Zuhörer. Diese Halligkeit ist nicht im Sinne eines möglichst natürlichen Klangergebnisses. Auf der anderen Seite ermöglicht es diese Verzögerung aber auch, dass der Richtungseindruck korrekt erhalten bleibt. Der Grund hierfür ist das Gesetz der ersten Wellenfront, welches sich wie folgt erklären lässt: Zwei Schallquellen sind in gleichem Abstand, jedoch an unterschiedlichen Orten um einen Zuhörer positioniert. Beide Schallquellen spielen das gleiche Signal ab, eine Quelle jedoch zeitlich verzögert. Dann wird der Zuhörer das Schallereignis bei der unverzögerten Quelle lokalisieren, auch wenn die verzögerte Quelle einen höheren Pegel hat. Für das ICC-System bedeutet dies, dass die Lautsprechersignale einen höheren Pegel haben können als der Direktschall, aber trotzdem der Sprecher als Schallquelle korrekt lokalisiert wird. Nach [55] können die höchsten Verstärkungen bei korrekter Sprecherlokalisierung bei einer Verzögerung von etwa 10 ms bis 15 ms erreicht werden.

Für den Sprecher selbst spielt die Verzögerung ebenfalls eine Rolle, da dieser seine eigene Stimme über das System wahrnimmt. Bei einer zu großen Verzögerung wird dieses Echo als störend empfunden. Um den Effekt zu minimieren sollte auch hier die Verzögerung einen Wert von ca. 10 ms bis 15 ms nicht überschreiten.

Um ein klanglich hochwertiges System zu realisieren, muss die Laufzeit demnach so gering wie möglich gehalten werden. Die Verzögerung kommt zum einen durch die Puffer der benötigten AD-/DA Wandler zustande. Um Rechenlast zu sparen, werden die Algorithmen oftmals im Frequenzbereich berechnet. Durch die dafür notwendige Blockverarbeitung wird eine zusätzliche Verzögerung ins System eingefügt, welche durch spezielle laufzeitoptimierte Filterbänke so gering wie möglich gehalten wird. In [75] wird beispielsweise eine *Overlap-Add-Filterbank* beschrieben, die zu diesem Zweck ein im Vergleich zum Analysefenster kürzeres Synthesefenster verwendet.

2.3.3 Unterdrückung der Hintergrundgeräusche

Zu den größten akustischen Herausforderungen neben der Rückkopplung gehört der hohe Hintergrundgeräuschpegel, welcher dem Sprachsignal überlagert ist und vom Mikrofon ebenfalls mit aufgenommen wird. Dieser ist insbesondere kritisch, als dass die Mikrofonpositionen aufgrund der Gegebenheiten im Fahrzeug oftmals weit vom Mund des Sprechers entfernt sind, wodurch das Signal-Rausch-Verhältnis (*Signal-to-Noise Ratio*, SNR) am Mikrofon zusätzlich beeinträchtigt wird. Ebenso sind die Lautsprecher meist nicht in unmittelbarer Ohrnähe des Zuhörers. Die daraus resultierende akustische Um-

gebung, in der ein ICC-System arbeitet, ist in [42] dargestellt. Der Abstand zwischen Sprechermund und Mikrofon kann durch in die Sicherheitsgurte integrierte Mikrofone reduziert werden. Damit ist im Vergleich zu konventionellen Einbaupositionen im Dach ein SNR-Gewinn von bis zu 10 dB erreichbar. Auch die Lautsprecher können beispielsweise durch Integration in die Kopfstütze in der Nähe der Zuhörerohren platziert werden [5]. Der Nachteil der Gurtmikrofone und der Kopfstützenlautsprecher sind die hohen Systemkosten.

Wie eingangs bereits erwähnt, setzt sich das Hintergrundgeräusch hauptsächlich aus Motorgeräusch, Windgeräuschen und Abrollgeräuschen der Reifen zusammen. In [29] werden diese unterschiedlichen Geräusche im Detail analysiert. Ein Ergebnis der Untersuchungen ist, dass der Großteil der Leistung des Motorgeräusches unterhalb 500 Hz liegt. Das Motorgeräusch hängt von der Drehzahl des Motors ab und kann sich dementsprechend z. B. bei Schaltvorgängen schnell ändern. Im Gegensatz dazu sind die Windgeräusche über einen längeren Zeitraum stationär, da sie hauptsächlich von der Geschwindigkeit des Fahrzeugs abhängen und sich dementsprechend maßgeblich beim Beschleunigen oder Abbremsen ändern. Das Spektrum der Windgeräusche ist breitbandiger als das des Motorgeräusches, allerdings ist auch hier der Großteil der Leistung unterhalb 1000 Hz konzentriert. Es wird weiter in derselben Quelle erläutert, dass der gleiche Frequenzbereich auch die meisten Anteile der Abrollgeräuschen der Reifen enthält. Das Abrollgeräusch ist ebenfalls stationär, kann sich jedoch z. B. bei einem Wechsel des Fahrbahnbelags schlagartig ändern.

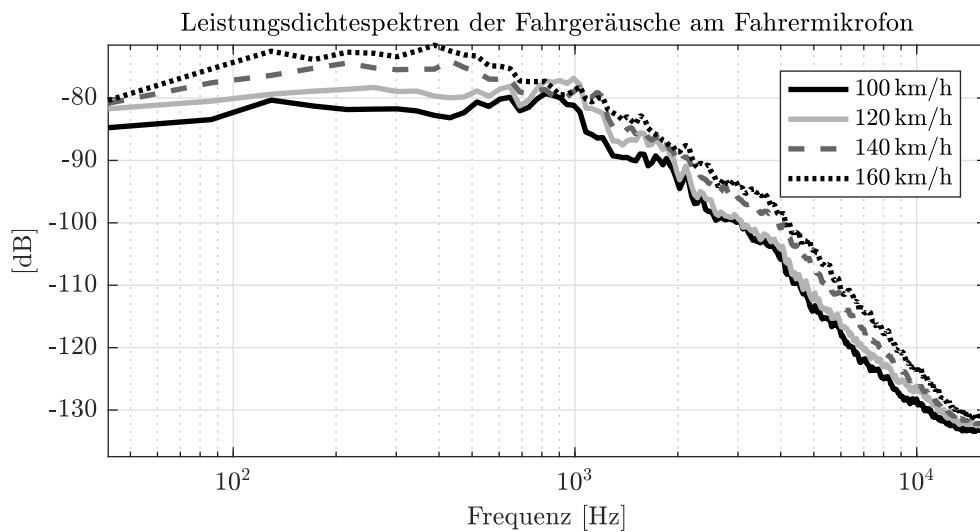


Abbildung 2.14: Leistungsdichtespektren der Fahrgeräusche, aufgenommen mit dem Fahrermikrofon bei verschiedenen Geschwindigkeiten. Die Lüftung im Fahrzeug war während der Messung ausgeschaltet.

In Abb. 2.14 sind die mittleren Leistungsdichtespektren (LDS) der Fahrgeräusche bei verschiedenen Geschwindigkeiten gezeigt. Die Geräusche wurden mit dem Fahrermikrofon im oben beschriebenen Demonstratorfahrzeug bei ausgeschalteter Lüftung aufge-

zeichnet. Man erkennt, dass mit zunehmender Geschwindigkeit auch der Geräuschpegel über den kompletten Frequenzbereich zunimmt. Insgesamt erkennt man zudem, dass die meiste Leistung im Frequenzbereich zwischen 100 Hz und 1000 Hz liegt, während die Leistung oberhalb 1000 Hz näherungsweise logarithmisch linear abfällt. Dies entspricht den oben erläuterten Ergebnissen aus [29].

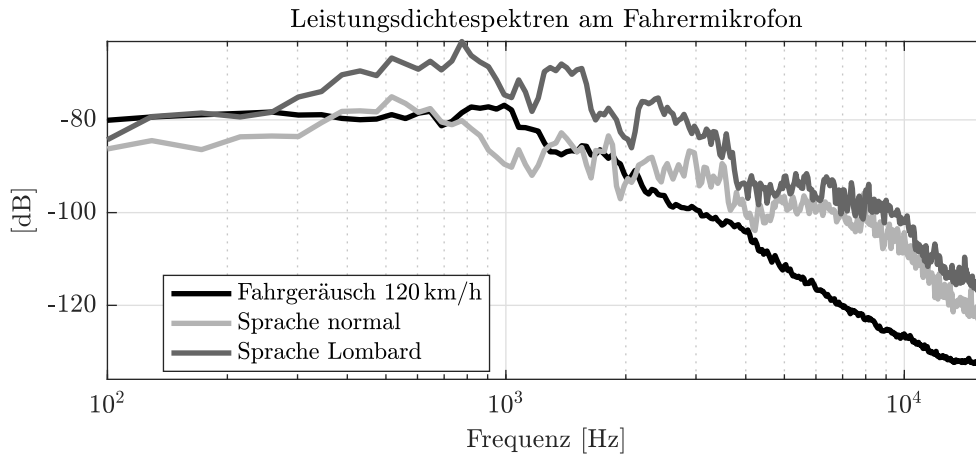


Abbildung 2.15: Leistungsdichtespektren von Sprache und Fahrgeräusch am Fahrermikrofon. Der Fahrer spricht entspannt mit normaler Lautstärke (hellgrau) und mit einem dem Fahrgeräusch angepassten, deutlich hörbaren Lombard-Effekt (dunkelgrau).

Wie stark das Fahrgeräusch der Sprache überlagert ist, zeigt Abb. 2.15. Die schwarze Kurve ist das mittlere LDS des Fahrgeräusches bei 120 km/h. Die beiden grauen Kurven sind die mittleren Leistungsdichtespektren des selben Sprechers. Das hellgraue Leistungsdichtespektrum erhält man, wenn der Sprecher in einer ruhigen Umgebung, entspannt und mit normaler Lautstärke spricht. Man erkennt, dass der Sprachpegel bis zu einer Frequenz von ca. 2000 Hz ungefähr dem Geräuschpegel entspricht, bzw. um 1000 Hz sogar deutlich darunter liegt. Erst oberhalb 2000 Hz liegen die Sprachanteile über dem Fahrgeräusch. Das A-gewichtete SNR liegt hier bei ca. -0,5 dBA. Spricht der Sprecher mit einer dem Hintergrundgeräusch angepassten Lautstärke, erhält man den dunkelgrauen Verlauf. In diesem Fall beträgt das A-gewichtete SNR etwa 12 dBA. Der Sprachpegel liegt über dem kompletten Frequenzbereich oberhalb des Hintergrundgeräusches. Allerdings ist der Lombard-Effekt hierbei deutlich wahrnehmbar, d. h. für den Sprecher ist es sehr anstrengend, diesen Pegel über einen längeren Zeitraum aufrecht zu halten. Bei den angegebenen Werten handelt es sich um Einzelmessungen. Für eine allgemeinere Betrachtung sei auf [3, S. 590 ff.] verwiesen. Dort wird angegeben, dass bei einem Schalldruckpegel des Geräusches von 62 dBA eine Unterhaltung mit normaler Stimme möglich ist, wenn die Gesprächspartner einen Abstand kleiner als 1,20 m zueinander haben. Nimmt der Geräuschpegel um ca. 12 dB zu, erfordert dies bereits eine sehr laute Stimme. Ist der Abstand größer, muss die Stimme entsprechend bereits bei geringeren Geräuschpegeln angehoben werden.

Die Messungen von Fahrgeräusch und Sprache zeigen, dass in einem ICC-System Maßnahmen zur Unterdrückung der Hintergrundgeräusche unumgänglich sind. Für Freisprechsysteme in Fahrzeugen existieren zahlreiche Verfahren zur Geräuschreduktion (vgl. z. B. [29, 43, 44, 13]). Da sich die Anforderungen von ICC-Systemen und Freisprecheinrichtungen an dieser Stelle nicht unterscheiden, kommen bei ICC-System meist ähnliche Verfahren zum Einsatz.

2.4 Zusammenfassung

In diesem Kapitel wurde das Demonstratorfahrzeug sowie die Randbedingungen im Fahrzeug in Form von Messungen der akustischen Eigenschaften vorgestellt. Hierbei wurden Erkenntnisse über die Totzeit und die Nachhallzeit des Fahrzeuginnenraumes gewonnen, welche für die spätere Parametrierung der Algorithmen eine wichtige Rolle spielen. Darüber hinaus zeigte sich, dass sich die Impulsantworten der Rückkopplungspfade, d. h. von den Lautsprechern zu den Mikrofonen, abhängig vom Beladungszustand und Bewegungen der Passagiere stark ändern können. Dies macht adaptive Algorithmen erforderlich, welche sich selbstständig an geänderte Begebenheiten anpassen. Messungen der Hintergrundgeräusche zeigten zudem, dass eine Kommunikation im Fahrzeug ohne ICC-System nur mit deutlich erhöhter Lautstärke der Stimme möglich ist.

Kapitel 3

Grundlagen der Rückkopplungsunterdrückung

Dieses Kapitel soll die Basis für das Verständnis der Arbeit schaffen. Im ersten Teil des Kapitels wird das Stabilitätskriterium für die elektro-akustische Schleife hergeleitet. Anschließend wird ein allgemeiner Überblick über den aktuellen Forschungsstand zum Thema Unterdrückung der elektro-akustischen Rückkopplung gegeben. Im darauffolgenden Abschnitt werden die für Kap. 4 benötigten mathematischen Grundlagen der adaptiven Rückkopplungskompensation gelegt. Das nachfolgend erläuterte Prinzip der Rückkopplungsreduktion wird als Grundlage für Kap. 5 benötigt. Am Ende dieses Kapitels wird die in dieser Arbeit verwendete Filterbankstruktur vorgestellt.

3.1 Stabilität rückgekoppelter Systeme

Das grundlegende Problem von Systemen, in denen eine elektro-akustische Rückkopplung auftritt, ist die Instabilität. Ein rückgekoppeltes System ist nur so lange stabil, wie das Stabilitätskriterium nicht verletzt wird. Übersteigt die Verstärkung die Stabilitätsgrenze, wird das System instabil und schwingt sich auf. Dies ist deutlich wahrnehmbar als unangenehmes, lautes Pfeifen. Aber auch bereits bei Verstärkungen nahe unterhalb der Stabilitätsgrenze sind störende Artefakte in Form von abklingenden tonalen Komponenten sowie Verzerrungen hörbar.

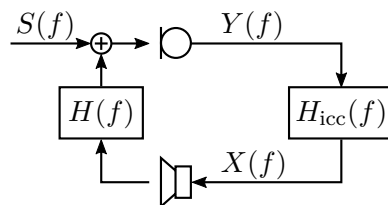


Abbildung 3.1: Schematisches Blockdiagramm einer geschlossenen elektro-akustischen Schleife. Das System besteht aus einem Mikrofon und einem Lautsprecher.

Zur Herleitung des Stabilitätskriteriums soll zunächst Abb. 3.1 betrachtet werden. In der Abbildung ist ein ICC-System gezeigt, welches aus einem Lautsprecher und einem Mikrofon besteht. Vereinfachend wird das ICC-System für die nachfolgenden Überlegungen als lineares, zeitinvariantes System modelliert. Die Kopplung zwischen Lautsprecher und Mikrofon wird durch die Übertragungsfunktion $H(f)$ des Raumes beschrieben.

$H_{\text{icc}}(f)$ bezeichnet die Übertragungsfunktion des ICC-Systems. Diese Übertragungsfunktion beinhaltet die zum Einsatz kommende Signalverarbeitung sowie die einstellbare Verstärkung des Systems. Sämtliche Signale sind hier im Frequenzbereich dargestellt, wobei f die kontinuierliche Frequenz bezeichnet.

Der Eingang des rückgekoppelten Systems ist das lokale Sprachsignal $S(f)$, der Ausgang das Lautsprechersignal $X(f)$. Damit ergibt sich für die resultierende Übertragungsfunktion des Gesamtsystems

$$H_{\text{res}}(f) = \frac{X(f)}{S(f)} = \frac{H_{\text{icc}}(f)}{1 - H_{\text{icc}}(f) \cdot H(f)}. \quad (3.1)$$

Strebt der Nenner von Gl. (3.1) bei einer Frequenz gegen null, wird $H_{\text{res}}(f)$ bei dieser Frequenz betragsmäßig groß. Dies ist immer dann der Fall, wenn die Schleifenverstärkung $H_s(f)$ gegen den Wert eins läuft, d. h.

$$H_s(f) = H_{\text{icc}}(f) \cdot H(f) \rightarrow 1. \quad (3.2)$$

Die dadurch hervorgerufene Überhöhung im Betrag der resultierenden Übertragungsfunktion $H_{\text{res}}(f)$ ist in Form einer Färbung des Lautsprechersignals wahrnehmbar. Je näher sich die Schleifenverstärkung bei einer kritischen Frequenz dem Wert eins nähert, desto deutlicher ist die Überhöhung in Form eines langsam abklingenden Tones hörbar. Übersteigt die Schleifenverstärkung den Wert eins, führt dies dazu, dass sich das System bei der entsprechenden Frequenz aufschwingt und damit instabil wird.

Die Schleifenverstärkung ist eine komplexwertige Funktion, deren Polarform

$$H_s(f) = |H_s(f)| \cdot e^{j \angle H_s(f)} \quad (3.3)$$

lautet. Aus Gl. (3.3) folgt das Stabilitätskriterium nach Nyquist [48]. Dieses besagt, dass das System instabil wird, wenn beide der nachfolgenden Bedingungen erfüllt sind

$$|H_s(f)| \geq 1 \quad (3.4)$$

$$e^{j \angle H_s(f)} = 1, \quad (3.5)$$

wobei der Phasenterm nur eins wird für ganzzahlige Vielfache von 2π , d. h.

$$\angle H_s(f) = l \cdot 2\pi, \quad l \in \mathbb{Z}. \quad (3.6)$$

Kritische Frequenzen sind hierbei vor allem Resonanzfrequenzen des Raumes. Bei diesen wird die erste Bedingung nach Gl. (3.4) bei zunehmender Verstärkung als erstes erreicht.

Da der Raumfrequenzgang $H(f)$ nicht verändert werden kann, kann die Schleifenverstärkung nur durch das ICC-System, d. h. $H_{\text{icc}}(f)$ beeinflusst werden. Wie bereits erwähnt, beinhaltet $H_{\text{icc}}(f)$ unter anderem die Verstärkung des Systems. Letztere wird als konstanter Faktor breitbandig auf alle Frequenzen angewandt. Die maximal mögliche Verstärkung, bei der das Stabilitätskriterium gerade noch nicht verletzt wird, wird als MSV (maximale stabile Verstärkung) bezeichnet. Bei Verstärkungen oberhalb der

MSV schwingt sich das System auf. Damit auch die durch die Rückkopplung verursachte Färbung des Spektrums nahe der Stabilitätsgrenze nicht wahrnehmbar ist, muss die eingestellte Verstärkung deutlich unterhalb der MSV liegen. Nach [37] haben Hörtests ergeben, dass für Sprache der Grenzwert, ab dem Artefakte hörbar werden, 5 dB unterhalb der MSV liegt.

Die MSV, die man ohne weitere Maßnahmen zur Unterdrückung der Rückkopplung erreichen kann, ist für ein ICC-System nicht ausreichend. Aus diesem Grund sind Maßnahmen zur Kompensation oder zur Reduktion der Rückkopplung notwendig.

3.2 Stand der Technik

Das oben beschriebene Problem der elektro-akustischen Rückkopplung tritt immer dann auf, wenn zwischen einem Mikrofon und einem bzw. mehreren Lautsprecher keine vollständige akustische Trennung herrscht. Neben ICC-Systemen kann dies bei verschiedensten technischen Anwendungen der Fall sein. Die prominentesten Beispiele hierfür sind z. B. Hörgeräte und Beschallungsanlagen. Entsprechend viel Forschung zur Unterdrückung der akustischen Rückkopplung wurde in diesen beiden Themengebieten betrieben. Ein umfassender Überblick über die existierenden Verfahren für die genannten Anwendungsgebiete wird in [70] gegeben. Die in dem Artikel vorgestellten Verfahren können in die vier nachfolgend beschriebenen Gruppen eingeteilt werden.

Bei der Phasenmodulation bzw. dem Frequenzversatz werden die Frequenzen des Mikrofonpektrums verschoben, bevor sie über den Lautsprecher verstärkt wiedergegeben werden. Da sich nun bei Überhöhungen des Raumfrequenzgangs die Signalleistung nicht mehr aufaddiert, wird ein Aufschwingen verhindert [58]. Dabei ist darauf zu achten, dass bei einem zu großen Frequenzversatz der Klang stark beeinträchtigt wird, während bei einem zu geringen Frequenzversatz die Stabilität nicht wesentlich verbessert wird. Effektiv eingesetzt werden kann dieses Verfahren vor allem dann, wenn der Zuhörer nur das in der Frequenz verschobene Signal zu hören bekommt, nicht aber den Direktschall. Im Fahrzeug ist dies nicht gewährleistet, da sich am Ohr des Zuhörers der Direktschall und das Lautsprechersignal überlagern. Da die Frequenzen des Lautsprechersignals geringfügig versetzt zu denen des Direktschalls sind, kommt es zu hörbaren Schwebungen. Haben z. B. Direktschall und Lautsprechersignal am Zuhörerohr die gleiche Amplitude, so wird das zu hörende Signal mit der Frequenz des Frequenzversatzes moduliert. Als alleiniges Verfahren zur Verbesserung der Stabilität kommt der Frequenzversatz im Fahrzeug demnach nicht in Frage.

Die zweite Gruppe der vorgestellten Verfahren dämpft entweder das komplette Spektrum oder einzelne Frequenzbänder, sobald ein Aufschwingen detektiert wird. So wird erreicht, dass das Stabilitätskriterium nicht verletzt wird. Zwar ist dies ein sehr effektiver Weg, ein Aufschwingen des Systems zu verhindern, allerdings kann damit die MSV nicht angehoben werden, d. h. höhere Pegel können nicht realisiert werden. Zu diesen Verfahren zählen beispielsweise die in [12, 23, 69] vorgestellten adaptiven Kerbfilter. Diese detektieren aufschwingende Sinussignale und setzen schmalbandige IIR-Kerbfilter (engl. *Infinite Impulse Response*, Filter mit unendlicher Impulsantwort) an die entspre-

chende Frequenz. Ein ähnliches Prinzip wird in [21] für ICC-Systeme beschrieben. Hier werden harmonische Signalanteile mittels spektraler Subtraktion unterdrückt. Um die Stabilität weiter zu verbessern, wird das Verfahren zusätzlich mit einem modulierten Frequenzversatz kombiniert.

Beamforming, realisiert mit Mikrofon- oder Lautsprecherarrays, verhindert ein Einkoppeln des Lautsprechersignals in das Mikrofon durch Aufprägen einer entsprechenden Richtcharakteristik [55]. Da im Fahrzeug jedoch zahlreiche Lautsprecher verbaut sind, müssen entsprechend viele Richtungen gedämpft werden. Hinzu kommt, dass bauraum- und kostenbedingt der Einsatz von großen Mikrofonarrays im Fahrzeug schwer realisierbar ist.

Die letzte Gruppe der in dem Artikel vorgestellten Verfahren befasst sich mit der Modellierung des Raumes, in dem das System betrieben wird. Gelingt es, das Übertragungsverhalten zwischen Lautsprecher und Mikrofon zu schätzen, kann die geschätzte Rückkopplung vom Mikrofon abgezogen werden. Verschiedene Verfahren aus dieser Gruppe wurden ebenfalls für ICC-Systeme intensiv untersucht und sind auch Schwerpunkt der vorliegenden Arbeit. Speziell bei ICC-Systemen werden bei Verfahren, die versuchen den Raum nachzubilden, die zwei Ansätze *Rückkopplungsreduktion* und *Rückkopplungskompensation* unterschieden.

Bei der modellbasierten Rückkopplungsreduktion nach [78] werden vorab die Impulsantworten zwischen den relevanten Lautsprechern und Mikrofonen im Fahrzeuginnenraum gemessen. Aus den gemessenen Impulsantworten wird ein Hallmodell extrahiert. Mittels dieses Hallmodells kann das Leistungsdichtespektrum der Rückkopplung und dessen Abklingverhalten am Mikrofon geschätzt werden. Der geschätzte Rückkopplungsanteil wird dann mit Hilfe eines Wiener-Filters spektral vom Mikrofonenspektrum abgezogen. Dieses Verfahren ist gut geeignet, um Hallanteile, die bereits nahe unterhalb der Stabilitätsgrenze hörbar werden, zu unterdrücken und damit den Klang des Systems zu verbessern. Sollten jedoch größere Verstärkungen notwendig sein, ist es mit diesem Verfahren nur in geringem Maße möglich, die Stabilitätsgrenze, bzw. die MSV, nach oben zu verschieben.

Eine komplette Auslöschung der Rückkopplung, um damit beliebige Verstärkungen realisieren zu können, ist nur möglich, wenn es gelingt die Impulsantwort des Rückkopplungspfades zu schätzen und das geschätzte Rückkopplungssignal im Zeitbereich phasenrichtig vom Mikrofonensignal abzuziehen. Dies ist das Ziel bei der adaptiven Rückkopplungskompensation. Das Prinzip ist ähnlich der Echokompensation, wie sie z. B. bei Freisprecheinrichtungen zum Einsatz kommt. Die Impulsantwort zwischen Lautsprecher und Mikrofon wird hier mit Hilfe eines adaptiven Filters geschätzt. Durch Faltung des Lautsprechersignals mit der geschätzten Impulsantwort erhält man das zu kompensierende Echosignal.

In [45, 50, 49] wird ein gewöhnliches Echokompensationsverfahren für die Rückkopplungskompensation in einem ICC-System verwendet. Die Impulsantwort wird im Zeitbereich mit einem NLMS-Algorithmus geschätzt. In [19] wird ein Stereo-Echokompensator, welcher zur Kompensation der Radio- bzw. Mediasignale eingesetzt wird, für die Kompensation der Rückkopplung verwendet. In dem Artikel liegt der Fokus auf einer möglichst geringen Laufzeit der Algorithmen. Das Problem bei der Verwendung herkömmli-

cher Echokompensatoren ist jedoch, dass adaptive Algorithmen nur dann zur korrekten Lösung konvergieren können, wenn das Anregungssignal (hier das Lautsprechersignal) und die lokale Störung (hier die lokale Sprache) orthogonal sind [73]. Im Falle eines ICC-Systems sind diese beiden Signale jedoch idealerweise identisch, weshalb die Konvergenz adaptiver Filter in rückgekoppelten Systemen begrenzt ist. In [51] wird dieser fundamentale Unterschied zwischen Echokompensation und Rückkopplungskompensation aufgezeigt.

Eine Möglichkeit, die Konvergenz zu verbessern ist es, die Signale zu dekorrelieren. In [77, 26] kommt hierfür der oben bereits erläuterte Frequenzversatz zum Einsatz. In [76, 74] wird gezeigt, dass bereits ein geringer Frequenzversatz im Sprachspektrum genügt, um die Signale ausreichend zu dekorrelieren, ohne dass das klangliche Ergebnis beeinträchtigt wird. Eine weitere Methode, die Signale zu dekorrelieren, ist lineare Prädiktion. In [53] ist beschrieben, wie lineare Prädiktion die Konvergenz des adaptiven Filters verbessern kann. Hierbei wird ein adaptives Filter als Schattenfilter parallel zum eigentlichen Kompensationsfilter eingesetzt. Die Spektren der Eingangssignale dieses Schattenfilters werden mittels linearer Prädiktion so gefiltert, dass sie weiß werden. Das Filter ist dabei für Sprachsignale ausgelegt. Damit kann das Schattenfilter gegen die korrekte Lösung konvergieren. Zur Kompensation der Rückkopplung werden die Koeffizienten des Schattenfilters in das Filter für die Rückkopplungskompensation kopiert. In [68] wird der Ansatz für tonale Signale wie z. B. Musik erweitert.

Zusätzlich zur linearen Prädiktion wird in [54, 67] die Schrittweite des adaptiven Filters variabel gesteuert. Das Ziel einer Schrittweitensteuerung ist es, sowohl die Konvergenzgeschwindigkeit als auch den Endabgleich des adaptiven Filters zu verbessern. Ein direkter Vergleich sowie eine Kombination der beiden Dekorrelationsmaßnahmen Frequenzversatz und lineare Prädiktion wird in [64, 65] diskutiert, wobei in diesen Artikeln die Schrittweite ebenfalls variabel gesteuert wird. In [66] wird diese Schrittweitensteuerung um einen Mechanismus zur Detektion von Korrelation ergänzt. Im Fall einer starken Korrelation der Eingangssignale wird hier die Schrittweite verringert. Umgekehrt wird bei geringer Korrelation oder bei Raumänderungen die Schrittweite erhöht.

Speziell in [64] werden zwei allgemeine Ansätze zur Steuerung der Schrittweite verglichen. Eine erste Möglichkeit ist es demnach, die Schrittweite nach [2] zu steuern. Da hier keinerlei Näherungen bei der Herleitung der optimalen Schrittweite getroffen werden, ist das Ergebnis auch für den Fall der korrelierten Signale bei der geschlossenen Schleife gültig. Das Problem ist allerdings, dass für die Schrittweitensteuerung die Leistung der lokalen Sprache bekannt sein muss, was aufwändige Schätzverfahren erforderlich macht. Mögliche Schätzverfahren werden ebenfalls in [64, 65] beschrieben. Eine zweite Möglichkeit zur Steuerung der Schrittweite wird in [46] für die akustische Echokompensation hergeleitet. Bei dieser Herleitung wird die für die Rückkopplungskompensation zunächst unzulässige Annahme gemacht, dass das Anregungssignal und die lokale Sprache unkorreliert sind. In Kap. 4 der vorliegenden Arbeit wird gezeigt, dass eine Abwandlung dieser Schrittweitensteuerung mit speziellen Erweiterungen auch für die Rückkopplungskompensation eingesetzt werden kann. Hierfür werden für die Adaption kurze Zeitpunkte im Nachhall genutzt, während derer die Signale unkorreliert sind. Das Resultat ist eine neuartige Schrittweitensteuerung, mit welcher ohne zusätzliche Dekorrelationsmaßnahmen

die Konvergenz eines Rückkopplungskompensators deutlich verbessert werden kann.

3.3 Prinzip der adaptiven Rückkopplungskompensation

Das in dieser Arbeit in Kap. 4 entwickelte Verfahren zur Verbesserung der Stabilität rückgekoppelter Systeme basiert auf der adaptiven Rückkopplungskompensation. Im Folgenden wird die grundlegende Idee der Rückkopplungskompensation kurz vorgestellt.

Wie bereits in Abschn. 3.2 erläutert, wird bei der adaptiven Rückkopplungskompensation versucht, die Übertragungsfunktion $H(f)$ zwischen Lautsprecher und Mikrofon mit einem adaptiven Filter zu schätzen. Kennt man die Übertragungsfunktion, kann man auch die am Mikrofon ankommende Rückkopplung kompensieren, indem man die geschätzte Rückkopplung vom Mikrofon subtrahiert. Dies ist in Abb. 3.2 dargestellt. Das

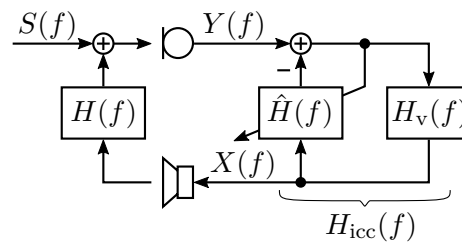


Abbildung 3.2: Schematisches Blockdiagramm eines rückgekoppelten Systems mit adaptiver Rückkopplungskompensation.

ICC-System besteht hier aus dem adaptiven Filter $\hat{H}(f)$ sowie dem Vorwärtspfad $H_v(f)$. $\hat{H}(f)$ ist die Schätzung der Übertragungsfunktion $H(f)$. Die Verstärkung des Systems ist Teil des Vorwärtspfades $H_v(f)$. Damit ergibt sich für die Übertragungsfunktion des ICC-Systems

$$H_{\text{icc}}(f) = \frac{X(f)}{Y(f)} = \frac{H_v(f)}{1 + H_v(f) \cdot \hat{H}(f)}, \quad (3.7)$$

wobei $Y(f)$ das Mikrofonsignal bezeichnet. Um die resultierende Übertragungsfunktion des Gesamtsystems zu erhalten, setzt man Gl. (3.7) in Gl. (3.1) ein. Dies führt auf

$$H_{\text{res}}(f) = \frac{X(f)}{S(f)} = \frac{H_v(f)}{1 - H_v(f) \cdot (H(f) - \hat{H}(f))}. \quad (3.8)$$

Das Ziel des adaptiven Filters ist es, eine perfekte Schätzung der realen Übertragungsfunktion des Raumes zu finden. Ist die Schätzung optimal, dann gilt

$$\hat{H}(f) = H(f). \quad (3.9)$$

Damit vereinfacht sich Gl. (3.8) zu

$$H_{\text{res}}(f) = H_v(f). \quad (3.10)$$

Die resultierende Übertragungsfunktion hängt dann also nur noch von der Übertragungsfunktion des Vorwärtspfades $H_v(f)$ ab, wodurch theoretisch beliebige Verstärkungen realisierbar sind. Sofern die Signalverarbeitung im Vorwärtspfad stabil ist, ist auch das Gesamtsystem stabil.

Während die Idee der adaptiven Rückkopplungskompensation anschaulich erklärbar ist, liegt die Schwierigkeit darin, ein adaptives Verfahren zu entwickeln, mit dem eine möglichst genaue Schätzung gefunden werden kann. Im nachfolgenden Abschnitt soll daher zunächst kurz die Theorie adaptiver Filter erläutert werden, bevor auf die speziellen Herausforderungen bei der adaptiven Rückkopplungskompensation eingegangen wird.

3.3.1 Theorie adaptiver Filter

Bei adaptiven Filtern handelt es sich um digitale Filter, die sich selbstständig an äußere Begebenheiten adaptieren. Dies wird realisiert, indem die Koeffizienten der Filter anhand vorgegebener Optimierungskriterien eingestellt werden. Ein tief reichendes Verständnis wird in [73, 31] gegeben. Im Folgenden werden nur die für die vorliegende Arbeit wesentlichen Aspekte erläutert.

Optimierungskriterium

Bei der adaptiven Rückkopplungskompensation wird ein adaptives Filter für die Systemidentifikation eingesetzt. Das vom adaptiven Filter zu identifizierende System ist dabei das System bestehend aus Lautsprechern, Fahrzeuginnenraum und Mikrofonen. Dieses System wird repräsentiert durch die Impulsantwort $\mathbf{h}(n)$

$$\mathbf{h}(n) = [h_0(n), h_1(n), \dots, h_{N_h-1}(n)]^T, \quad (3.11)$$

wobei N_h die Länge der Impulsantwort in Abtastwerten ist und n den diskreten Zeitindex bezeichnet. Das adaptive Filter beinhaltet die geschätzte Impulsantwort $\hat{\mathbf{h}}(n)$

$$\hat{\mathbf{h}}(n) = [\hat{h}_0(n), \hat{h}_1(n), \dots, \hat{h}_{N_h-1}(n)]^T. \quad (3.12)$$

Abb. 3.3 zeigt das Blockdiagramm des adaptiven Filters zur Systemidentifikation im zeitdiskreten Bereich. Das Ziel ist es, dass das Fehlersignal $e(n)$ nur noch die lokale

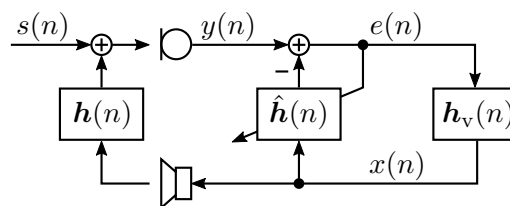


Abbildung 3.3: Adaptives Filter bei der Systemidentifikation.

Sprache $s(n)$, aber keine Rückkopplungsanteile mehr enthält. Nach Verarbeitung mit

dem Filter im Vorwärtspfad $\mathbf{h}_v(n)$ wird das Fehlersignal als Lautsprecher-signal $x(n)$ wieder ausgegeben. Mit der Vektorschreibweise

$$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-N_h+1)]^T \quad (3.13)$$

berechnet sich der Fehler zu

$$e(n) = s(n) + \mathbf{x}^T(n) \cdot \mathbf{h}(n) - \mathbf{x}^T(n) \cdot \hat{\mathbf{h}}(n) \quad (3.14)$$

$$= s(n) + \mathbf{x}^T(n) \cdot (\mathbf{h}(n) - \hat{\mathbf{h}}(n)). \quad (3.15)$$

Äquivalent zu Gl. (3.9) zeigt sich auch in Gl. (3.15), dass das Fehlersignal $e(n)$ genau dann der lokalen Sprache $s(n)$ entspricht, wenn das adaptive Filter der realen Impulsantwort entspricht. Für die optimalen Koeffizienten gilt also

$$\hat{\mathbf{h}}_{\text{opt}}(n) = \mathbf{h}(n). \quad (3.16)$$

Klassischerweise wird in der adaptiven Filter-Theorie zur Berechnung der optimalen Koeffizienten der mittlere quadratische Fehler minimiert [73]. Den mittleren quadratischen Fehler erhält man durch Quadrieren von Gl. (3.15) und anschließender Erwartungswertbildung $E\{\cdot\}$

$$\begin{aligned} E\{e^2(n)\} &= E\left\{\left(s(n) + \mathbf{x}^T(n)\mathbf{h}_\Delta(n)\right)^2\right\} \\ &= E\left\{s^2(n) + 2s(n)\mathbf{x}^T(n)\mathbf{h}_\Delta(n) + \left(\mathbf{x}^T(n)\mathbf{h}_\Delta(n)\right)^2\right\} \\ &= E\left\{s^2(n) + 2s(n)\mathbf{x}^T(n)\mathbf{h}_\Delta(n) + \mathbf{h}_\Delta^T(n)\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{h}_\Delta(n)\right\}. \end{aligned} \quad (3.17)$$

Zur besseren Lesbarkeit wurde folgende Substitution verwendet

$$\mathbf{h}_\Delta(n) = \mathbf{h}(n) - \hat{\mathbf{h}}(n). \quad (3.18)$$

Es wird im Folgenden angenommen, dass $\mathbf{x}(n)$ und $\mathbf{h}_\Delta(n)$ unkorreliert sind. Bezeichnet dann man mit $\mathbf{R}_{xx}(n)$ die Autokorrelationsmatrix des Lautsprecher- bzw. Anregungssignals

$$\mathbf{R}_{xx}(n) = E\{\mathbf{x}(n)\mathbf{x}^T(n)\} \quad (3.19)$$

und mit $\mathbf{r}_{sx}(n)$ den Kreuzkorrelationsvektor zwischen lokaler Sprache und Lautsprecher-signal

$$\mathbf{r}_{sx}(n) = E\{s(n)\mathbf{x}(n)\}, \quad (3.20)$$

so lässt sich Gl. (3.17) schreiben als

$$E\{e^2(n)\} = E\{s^2(n)\} + 2\mathbf{r}_{sx}^T(n)\mathbf{h}_\Delta(n) + \mathbf{h}_\Delta^T(n)\mathbf{R}_{xx}(n)\mathbf{h}_\Delta(n). \quad (3.21)$$

Gl. (3.21) ist die Kostenfunktion des Optimierungsproblems. Sie beschreibt eine quadratische Funktion, die von den Koeffizienten $\hat{\mathbf{h}}(n)$ abhängt. Weil sie quadratisch ist, hat

die Kostenfunktion genau ein globales Minimum. Dieses Minimum findet man an den Nullstellen der Ableitung von Gl. (3.21) nach $\hat{\mathbf{h}}(n)$. Da es sich bei der Kostenfunktion um eine mehrdimensionale Funktion von $\hat{\mathbf{h}}(n)$ handelt, ist die Ableitung durch den Gradienten $\nabla_{\hat{\mathbf{h}}(n)}\{\cdot\}$

$$\nabla_{\hat{\mathbf{h}}(n)}\left\{\mathbb{E}\left\{e^2(n)\right\}\right\} = \frac{\partial\mathbb{E}\left\{e^2(n)\right\}}{\partial\hat{\mathbf{h}}(n)} = -2\mathbf{r}_{sx}(n) - 2\mathbf{R}_{xx}(n)\left(\mathbf{h}(n) - \hat{\mathbf{h}}(n)\right) \quad (3.22)$$

gegeben. Nullsetzen und Auflösen von Gl. (3.22) führt schließlich auf die Koeffizienten

$$\hat{\mathbf{h}}_{\text{opt}}(n) = \mathbf{h}(n) + \mathbf{R}_{xx}^{-1}(n)\mathbf{r}_{sx}(n). \quad (3.23)$$

Die Koeffizienten $\hat{\mathbf{h}}_{\text{opt}}(n)$ sind also die Koeffizienten, welche die Kostenfunktion nach Gl. (3.21) minimieren. Es fällt auf, dass diese Koeffizienten nur unter einer bestimmten Bedingung mit dem gewünschten Resultat nach Gl. (3.16) übereinstimmen. Diese Bedingung ist, dass die lokale Sprache $s(n)$ und das Lautsprechersignal $x(n)$ orthogonal zueinander sind. Ist dies der Fall, gilt $\mathbf{r}_{sx}(n) = \mathbf{0}_{N_h}$, wobei $\mathbf{0}_{N_h}$ ein Nullvektor der Länge N_h ist, und damit $\hat{\mathbf{h}}_{\text{opt}}(n) = \mathbf{h}(n)$. Im Fall des rückgekoppelten ICC-Systems soll jedoch der Lautsprecher idealerweise die lokale Sprache unverändert wiedergeben. D. h. es besteht eine starke lineare Abhängigkeit zwischen den beiden Folgen $x(n)$ und $s(n)$. Dies ist der Grund, weshalb Standardverfahren adaptiver Filter in einer geschlossenen elektro-akustischen Schleife nicht gegen die gewünschte Lösung nach Gl. (3.16) konvergieren können, sondern stattdessen gegen die in Gl. (3.23) angegebene Lösung konvergieren (vgl. auch [51]).

Wie bereits in Abschn. 3.2 beschrieben wurde, ist ein naheliegender Ansatz, die Signale im Vorwärtspfad $\mathbf{h}_v(n)$ zu dekorrelieren, um so eine bessere Konvergenz zu erreichen. Dabei tritt jedoch die Schwierigkeit auf, dass bei einer zu starken Dekorrelation der Klang des Lautsprechersignals stark beeinträchtigt wird, während bei einer schwachen Dekorrelation die Konvergenz des Filters nicht ausreichend verbessert wird. In der vorliegenden Arbeit liegt der Fokus daher auf einem anderen Ansatz. Dieser nutzt für die Adaption kurze Zeitpunkte im Nachhall aus, während derer die Signale unkorreliert sind.

Normalized-Least-Mean-Square-Algorithmus

Zur Berechnung adaptiver Filter existieren zahlreiche verschiedene Algorithmen. Allen ist gemeinsam, dass sie mittels numerischer Optimierungsverfahren versuchen, das Minimum der Kostenfunktion nach Gl. (3.21) zu finden. Sind die Signale $x(n)$ und $s(n)$ orthogonal, entspricht das gefundene Minimum der gewünschten Lösung $\hat{\mathbf{h}}(n) = \mathbf{h}(n)$.

Das bekannteste Verfahren ist der in den 1960er Jahren von Bernard Widrow und Marcian E. Hoff entwickelte *Least-Mean-Square* (LMS)-Algorithmus [72]. Aufgrund der geringen Rechenkomplexität kommt der Algorithmus in verschiedensten Bereichen der digitalen Signalverarbeitung zum Einsatz. Der Algorithmus gehört zur Gruppe der Gradientenverfahren. Bei diesen Verfahren werden die Filterkoeffizienten iterativ aktualisiert, indem man sich mit einer bestimmten Schrittweite immer in Richtung des steilsten Abstiegs entlang der Kostenfunktion bewegt. Die Richtung des steilsten Abstiegs wird

durch den negativen Gradienten der Kostenfunktion festgelegt. Ausgehend von einem Startwert $\hat{\mathbf{h}}(n=0)$ gilt damit bei den Gradientenverfahren für die rekursive Aktualisierung der Koeffizienten

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{1}{2} \cdot \alpha(n) \cdot \left(-\nabla_{\hat{\mathbf{h}}(n)} \left\{ \mathbb{E} \left\{ e^2(n) \right\} \right\} \right). \quad (3.24)$$

Hierbei ist $\alpha(n)$ ein Parameter, mit welchem die Schrittweite der Aktualisierungsschritte eingestellt werden kann. Der Faktor $1/2$ halbiert die Schrittweite $\alpha(n)$ und wird nur aus „kosmetischen“ Gründen eingeführt [31].

Betrachtet man Gl. (3.22), erkennt man, dass der Gradient nicht direkt berechnet werden kann, da sowohl die reale Impulsantwort $\mathbf{h}(n)$ als auch der Autokorrelationsvektor $\mathbf{r}_{sx}(n)$ unbekannt sind. Beim LMS-Algorithmus wird daher der Gradient angenähert, indem anstelle des Erwartungswertes des quadratischen Fehlers der Momentanwert verwendet wird

$$\nabla_{\hat{\mathbf{h}}(n)} \left\{ \mathbb{E} \left\{ e^2(n) \right\} \right\} \approx \nabla_{\hat{\mathbf{h}}(n)} \left\{ e^2(n) \right\} = -2e(n)\mathbf{x}(n). \quad (3.25)$$

Setzt man den geschätzten Gradienten nach Gl. (3.25) in Gl. (3.24) ein, erhält man den LMS-Algorithmus

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \alpha(n) \cdot e(n)\mathbf{x}(n). \quad (3.26)$$

Ein Problem des LMS-Algorithmus ist die Abhängigkeit des geschätzten Gradienten von der Leistung des Anregungssignals. Ist die Anregung durch das Lautsprechersignal klein, wird auch der Gradient klein, während bei einer starken Anregung der Gradient entsprechend groß ist. Vor allem bei Sprache schwankt die Kurzzeitleistung stark, was zur Folge hat, dass der Gradient in gleichem Maße schwankt. Allgemein kann man sagen, dass das Minimum der Kostenfunktion umso präziser gefunden wird, je kleiner die Schritte sind, mit denen man sich auf der Kostenfunktion in Richtung des Minimums bewegt. Der Grund ist, dass bei kleinen Aktualisierungsschritten ein einzelner Schritt in die falsche Richtung weniger stark ins Gewicht fällt. Der Nachteil ist die langsamere Konvergenz. Umgekehrt wird der LMS-Algorithmus bei zu großen Aktualisierungsschritten instabil, d. h. er divergiert.

Aus diesem Grund wird beim *Normalized-Least-Mean-Square* (NLMS)-Algorithmus die Aktualisierung der Koeffizienten zwischen zwei Iterationsschritten minimal gehalten. Dies wird realisiert, indem der Gradient, bzw. die Schrittweite, auf die quadratische euklidische Norm des Anregungssignals normiert wird. Eine detaillierte Herleitung des NLMS-Algorithmus ist z. B. in [31] gegeben. Es folgt für die Aktualisierung der Filterkoeffizienten beim NLMS-Algorithmus

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \alpha(n) \cdot \frac{e(n)\mathbf{x}(n)}{\|\mathbf{x}(n)\|^2}. \quad (3.27)$$

Hierbei bezeichnet $\|\cdot\|$ die euklidische Norm eines Vektors, d. h.

$$\|\mathbf{x}(n)\|^2 = \sum_{i=0}^{N_h-1} |x(n-i)|^2. \quad (3.28)$$

3.3.2 Implementierung adaptiver Filter im Frequenzbereich

Für die Berechnungen adaptiver Filter sind im Zeitbereich Faltungen mit der geschätzten Impulsantwort notwendig. Da die zu schätzenden Impulsantworten im Fahrzeug lang sind, sind diese Faltungen rechenaufwändig. Die Rechenleistung der im Fahrzeug zum Einsatz kommenden Prozessoren ist aus Kosten- und Platzgründen begrenzt. Zur Verringerung der Rechenlast wird aus diesem Grund die schnelle Faltung im Frequenzbereich angewendet. Dazu werden die Signale mit Filterbänken blockweise in den Frequenzbereich transformiert und dann in einzelnen Teilbändern verarbeitet. Speziell für die adaptiven Filter ergibt sich durch die frequenzselektive Normierung des Gradienten im Teilband zudem der Vorteil einer besseren Konvergenz [31].

Schnelle Faltung mit *Overlap-Save-Filterbank*

Im Folgenden wird die schnelle Faltung mit einer *Overlap-Save-Filterbank* beschrieben. Das Verfahren wird in ähnlicher Form auch z. B. in [36] hergeleitet. Der Abschnitt dient hauptsächlich dazu, mit der in dieser Arbeit verwendeten Schreibweise vertraut zu machen.

Ein zeitdiskretes Signal $x(n)$ soll im Zeitbereich mit der Impulsantwort $\mathbf{h}(n)$ aperiodisch gefaltet werden. Die Impulsantwort sei zunächst zeitinvariant, d. h. es gilt

$$\mathbf{h}(n) = \mathbf{h}(n + 1) = \mathbf{h}. \quad (3.29)$$

Das gefilterte Ausgangssignal $r(n)$ berechnet sich dann über die Faltungssumme wie folgt

$$r(n) = \sum_{i=0}^{N_h-1} x(n-i) \cdot h_i. \quad (3.30)$$

Diese Operation nimmt bei langen Impulsantworten viel Rechenleistung in Anspruch. Die Rechenleistung kann wesentlich verringert werden, wenn stattdessen die Signale blockweise mittels diskreter Fouriertransformation (DFT) in den Frequenzbereich transformiert werden und dort eine Multiplikation der Spektren durchgeführt wird. Dabei ist darauf zu achten, dass die Multiplikation im Frequenzbereich einer *zyklischen* Faltung im Zeitbereich entspricht, weshalb sogenannte Projektionen notwendig sind, um die gewünschte aperiodische Faltung zu erhalten.

Dies soll hier am Beispiel einer *Overlap-Save-Filterbank* gezeigt werden. Die Blockgröße entspricht der DFT-Länge N . Hinsichtlich der Rechenkomplexität ist die effizienteste Implementierung der *Overlap-Save-Struktur* gegeben, wenn sich die Blöcke halb überlappen [31]. Dies soll auch in der vorliegenden Arbeit verwendet werden, d. h. der Rahmenversatz ist $L = N/2$. Der Nachteil der blockweisen Verarbeitung ist eine Verzögerung des Ausgangssignals, abhängig von der Blockgröße. Bei der *Overlap-Save-Methode* entspricht die Verzögerung dem Rahmenversatz L . Um diese Verzögerung möglichst klein zu halten, muss demnach auch die Blockgröße, bzw. der Rahmenversatz klein gehalten werden. Bei langen Impulsantworten muss dafür die Impulsantwort in M kürzere Blöcke zerteilt werden. Jeder dieser Blöcke wird dann mit dem zeitlich korrespondierenden Eingangsblock im Frequenzbereich multipliziert. Zur Unterdrückung der Fehler durch die

zyklische Faltung wird bei der hier vorgestellten *Overlap-Save*-Methode die rechte Hälfte eines Impulsantwortblocks auf null gesetzt. Das Prinzip der blockweisen Verarbeitung ist in Abb. 3.4 dargestellt.

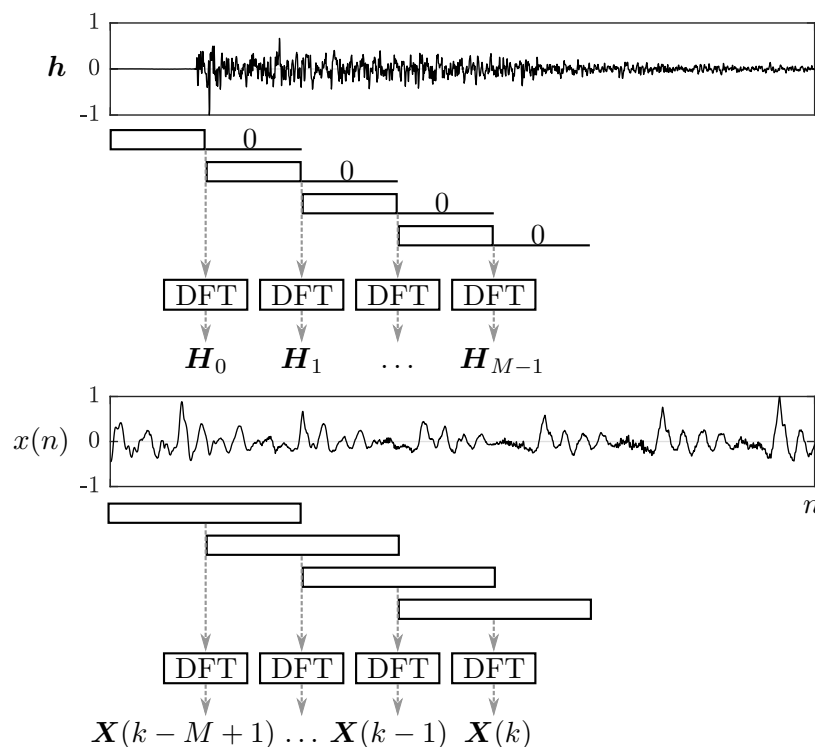


Abbildung 3.4: Blockweise Verarbeitung bei der schnellen Faltung mit einer *Overlap-Save*-Filterbank mit halber Überlappung der Blöcke.

Die Vektoren \mathbf{H}_m , mit $m = 0, \dots, M - 1$, beinhalten die diskreten DFT-Spektren des jeweiligen Blocks

$$\begin{aligned} \mathbf{H}_m &= \text{DFT} \left\{ \left[h_{mL}, h_{mL+1}, \dots, h_{(m+1)L-1}, \mathbf{0}_L \right]^T \right\} \\ &= [H_m(0), H_m(1), \dots, H_m(N-1)]^T. \end{aligned} \quad (3.31)$$

Diese Vektoren werden fortan *Teilfilter* genannt. $\mathbf{0}_L$ bezeichnet einen Nullvektor der Länge L . Der diskrete Frequenzindex ist $\mu = 0, \dots, N - 1$. Analog dazu beinhalten die Vektoren $\mathbf{X}(k)$ die Spektren des k -ten Blockes des Eingangssignals

$$\begin{aligned} \mathbf{X}(k) &= \text{DFT} \left\{ [x(n-N+1), \dots, x(n-1), x(n)]^T \right\} \\ &= [X(0, k), X(1, k), \dots, X(N-1, k)]^T, \end{aligned} \quad (3.32)$$

wobei $k = n/L$ den unterabgetasteten Zeitindex der Blockverarbeitung bezeichnet.

Das Spektrum des Ausgangssignals $\mathbf{R}(k) = [R(0, k), R(1, k), \dots, R(N-1, k)]^T$ erhält man, indem man die Spektren $\mathbf{X}(k-m)$ elementweise mit den zugehörigen Blöcken

\mathbf{H}_m multipliziert und schließlich über m aufsummiert. *Elementweise* meint hierbei die einzelnen Teilbänder μ . Für das μ -te Teilband des Ausgangssignals gilt also

$$R(\mu, k) = \sum_{m=0}^{M-1} X(\mu, k - m) \cdot H_m(\mu). \quad (3.33)$$

Man erkennt, dass Gl.(3.33) einer Faltung im Zeitbereich entspricht. $X(\mu, k)$ ist der zeitliche Verlauf des μ -ten Teilbands. Die zugehörige Zeitbasis ist der unterabgetastete Blockindex k . Der Verlauf jedes Teilbands μ wird also mit den M Filterkoeffizienten $H_m(\mu)$ gefaltet. Üblicherweise beschreiben die Koeffizienten einer Impulsantwort die zeitliche Folge eines Zeitsignals. Analog dazu werden die Filterkoeffizienten des μ -ten Teilbands fortan *Teilbandimpulsantworten* genannt und sind im Vektor

$$\mathbf{H}(\mu) = [H_0(\mu), H_1(\mu), \dots, H_{M-1}(\mu)]^T \quad (3.34)$$

zusammengefasst. Auch die vergangenen Werte eines Teilbands des Eingangssignals können in Vektorschreibweise ausgedrückt werden

$$\mathbf{X}(\mu, k) = [X(\mu, k), X(\mu, k - 1), \dots, X(\mu, k - M + 1)]^T. \quad (3.35)$$

Damit lautet die Vektorschreibweise von Gl.(3.33)

$$R(\mu, k) = \mathbf{X}^T(\mu, k) \cdot \mathbf{H}(\mu). \quad (3.36)$$

Das Zeitbereichssignal $r(n)$ erhält man schließlich durch inverse DFT (IDFT)

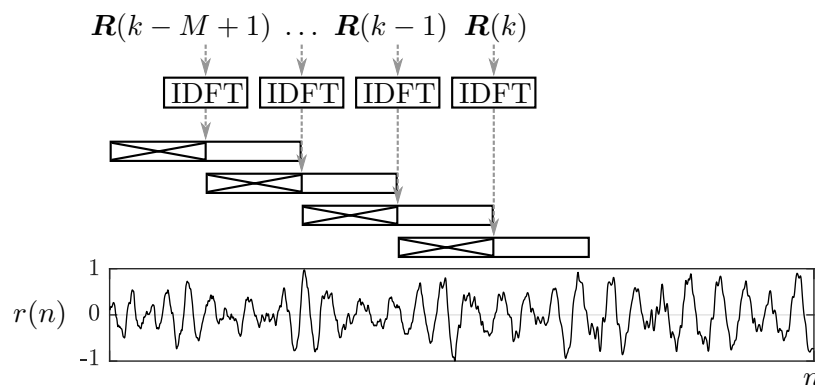
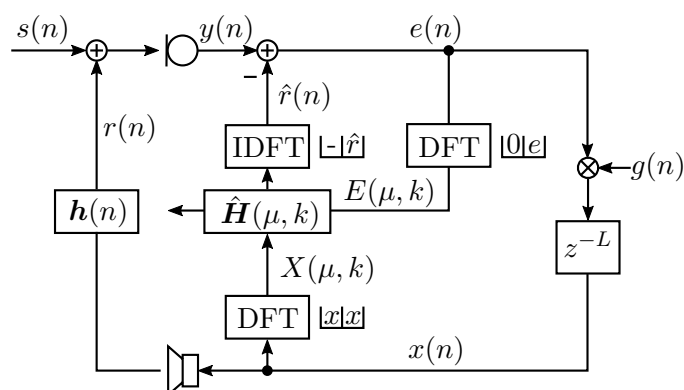
$$\begin{aligned} \mathbf{r}(n) &= \text{IDFT}\{\mathbf{R}(k)\} \\ &= [r(n - N + 1), \dots, r(n - 1), r(n)]^T. \end{aligned} \quad (3.37)$$

Dabei ist zu beachten, dass aufgrund der zyklischen Überfaltung die ersten L Werte des Zeitbereichssignals nicht mit dem gewünschten aperiodisch gefalteten Signal übereinstimmen. Bei der *Overlap-Save*-Methode werden diese daher verworfen und nur die rechte Hälfte der Blöcke wird aneinander gereiht. Dies ist schematisch in Abb. 3.5 dargestellt.

Frequenzbereichsimplementierung des NLMS-Algorithmus

Mit Hilfe der oben erläuterten blockweisen Signalverarbeitung können auch adaptive Filter im Teilband realisiert werden [62, 61]. Während adaptive Filter bei der Implementierung im Zeitbereich die tatsächliche, breitbandige Impulsantwort des Raumes schätzen, wird bei der blockbasierten Verarbeitung für jedes Teilband die jeweilige Teilbandimpulsantwort geschätzt.

In Abb. 3.6 ist das Blockdiagramm der adaptiven Rückkopplungskompensation gezeigt. Für die schnelle Faltung kommt die in Abschn. 3.3.2 erläuterte *Overlap-Save*-Filterbank mit halb überlappenden Blöcken zum Einsatz. Die dabei verarbeiteten Zeitbereichsblöcke sind schematisch rechts neben den DFT/IDFT Blöcken dargestellt.


 Abbildung 3.5: Zusammensetzen des Ausgangssignals bei der *Overlap-Save*-Methode.

 Abbildung 3.6: Blockdiagramm der Implementierung eines adaptiven Filters im Teilband. Für die Transformation in den Frequenzbereich wird eine *Overlap-Save*-Filterbank verwendet.

Beim Lautsprechersignal $x(n)$ wird ein kompletter Block bestehend aus den N letzten Abtastwerten in den Frequenzbereich transformiert (vgl. Gl. (3.32)). Beim Fehlersignal müssen zur Unterdrückung der Fehler durch die zyklische Faltung die ersten L Werte des Fehlerblocks auf null gesetzt werden, d. h.

$$\begin{aligned} \mathbf{E}(k) &= \text{DFT} \left\{ [\mathbf{0}_L, e(n-L+1), \dots, e(n)]^T \right\} \\ &= [E(0, k), E(1, k), \dots, E(N-1, k)]^T. \end{aligned} \quad (3.38)$$

Die geschätzte Rückkopplung $\hat{r}(n)$ ist das Ergebnis der schnellen Faltung des Lautsprechersignals $\mathbf{X}(\mu, k)$ mit dem adaptiven Filter $\hat{\mathbf{H}}(\mu, k)$. Wie oben erläutert, werden bei der *Overlap-Save*-Methode die ersten L Werte dieses Blocks verworfen. Das adaptive Filter besteht aus den N geschätzten Teilbandimpulsantworten

$$\hat{\mathbf{H}}(\mu, k) = [\hat{H}_0(\mu, k), \hat{H}_1(\mu, k), \dots, \hat{H}_{M-1}(\mu, k)]^T. \quad (3.39)$$

Die zu schätzende reale Impulsantwort des Raumes $\mathbf{h}(n)$ ist zeitvariant. Der Grund hierfür ist, dass sich im realen System die Impulsantwort ändert, wenn sich z. B. die Insassen im Fahrzeug bewegen oder ein Fenster geöffnet wird (vgl. Kap. 2). Die reale Rückkopplung $r(n)$ ergibt sich durch Faltung von $x(n)$ und $\mathbf{h}(n)$. Aus der Überlagerung von Rückkopplung $r(n)$ und lokaler Sprache $s(n)$ ergibt sich das Mikrofonsignal $y(n)$.

Im Vorwärtspfad des ICC-Systems befinden sich die einstellbare Systemverstärkung $g(n)$ und eine Verzögerung z^{-L} um L Abtastwerte. Wie in Abschn. 3.3.2 erläutert, wird diese Verzögerung systembedingt durch die blockbasierte Verarbeitung verursacht.

Das adaptive Filter wird mit dem NLMS-Algorithmus implementiert. Die rekursive Aktualisierung des adaptiven Filters erfolgt ähnlich zur Zeitbereichsimplementierung nach Gl. (3.27). Die Unterschiede sind zum einen, dass die Signale im Teilband komplexwertig sind und zum anderen, dass zur Unterdrückung der zyklischen Faltungseffekte Projektionen durchgeführt werden müssen. Zur besseren Lesbarkeit wird im Folgenden die geschätzte Impulsantwort im Teilband μ vor der Projektion mit $\hat{\mathbf{H}}_p(\mu, k)$ bezeichnet, nach der Projektion mit $\hat{\mathbf{H}}(\mu, k)$. Mit dieser Schreibweise ergibt sich für den NLMS im Teilband

$$\hat{\mathbf{H}}_p(\mu, k+1) = \hat{\mathbf{H}}_p(\mu, k) + \alpha(\mu, k) \cdot \frac{E(\mu, k) \cdot \mathbf{X}^*(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}. \quad (3.40)$$

Dabei ist $\alpha(\mu, k)$ die frequenzselektive Schrittweite.

Für die Projektionen müssen die Elemente des Vektors $\hat{\mathbf{H}}_p(\mu, k)$ umsortiert werden. Im Vektor $\hat{\mathbf{H}}_p(\mu, k)$ sind die M Koeffizienten der Teilbandimpulsantwort im Teilband μ zum Zeitpunkt k zusammengefasst. Sortiert man diese Elemente über die Frequenz jedes Teilfilters, erhält man M Spektren

$$\hat{\mathbf{H}}_{p,m}(k) = \left[\hat{H}_{p,m}(0, k), \hat{H}_{p,m}(1, k), \dots, \hat{H}_{p,m}(N-1, k) \right]^T. \quad (3.41)$$

Zur Unterdrückung der Faltungsfehler durch die zyklische Faltung, müssen diese Spektren in den Zeitbereich transformiert werden. Dort wird die rechte Hälfte des Zeitsignalkvektors auf null gesetzt. Der resultierende Vektor wird zurück in den Teilbandbereich transformiert. Dies wird durch folgende Gleichung ausgedrückt

$$\hat{\mathbf{H}}_m(k) = \text{DFT} \left\{ \text{IDFT} \left\{ \hat{\mathbf{H}}_{p,m}(k) \right\} \circ [\mathbf{1}_L, \mathbf{0}_L]^T \right\}, \quad (3.42)$$

wobei mit \circ eine elementweise Vektormultiplikation gemeint ist und $\mathbf{1}_L$ einen Vektor der Länge L bezeichnet, dessen Elemente alle den Wert eins haben. Erneut erhält man durch Umsortieren aus den Teilfiltern $\hat{\mathbf{H}}_m(k)$ die Teilbandimpulsantworten $\hat{\mathbf{H}}(\mu, k)$.

Neben der optimierten Rechenleistung hat die Implementierung adaptiver Filter im Teilband einen weiteren Vorteil. Beim NLMS im Zeitbereich (Gl. (3.27)) wird der Gradient auf die Gesamtleistung des Anregungssignals normiert. Ist die Leistung des Anregungssignals nur auf einzelne Frequenzbereiche konzentriert, wird auch der Gradient von diesen Bändern dominiert. Da der Adaptionsprozess hier breitbandig ist, ist die Normierung des Gradienten für Frequenzbänder mit wenig Anregungssignal nicht optimal. In Gl. (3.40) erkennt man, dass im Gegensatz dazu bei der Implementierung im Teilband der Gradient frequenzselektiv berechnet wird. D. h. auch die Leistungsnormierung wird

für jedes Teilband separat durchgeführt. Dadurch ist der Gradient für jedes Teilband optimal, was dem Fall einer weißen Anregung entspricht. Da weiße Prozesse unkorreliert sind, konvergieren adaptive Filter mit weißen Anregungssignalen am schnellsten [55]. Wie man jedoch ebenfalls in Gl. (3.40) erkennt, wird die Leistung nur aus den letzten M Werten geschätzt, die Normierung ist also nicht exakt. Man spricht bei der frequenzselektiven Normierung des Gradienten von einer Pseudo-Orthogonalisierung.

3.4 Prinzip der Rückkopplungsreduktion mit Wiener-Filter

Neben der im vorangegangenen Abschnitt vorgestellten Rückkopplungskompensation stellt ein weiteres Verfahren zur Unterdrückung der Rückkopplung die Rückkopplungsreduktion mittels Wiener-Filter dar. Das Prinzip basiert auf Verfahren zur Geräuschreduktion, wie z. B. in [4, 16, 17] beschrieben, wobei anstelle des geschätzten Geräusches die geschätzte Rückkopplung in das Wiener-Filter eingeht.

3.4.1 Herleitung des Wiener-Filters

Zur Herleitung des Wiener-Filters soll zunächst das Blockdiagramm in Abb. 3.7 betrachtet werden. Sämtliche Signale sind im Zeitbereich dargestellt. Um der Terminologie der

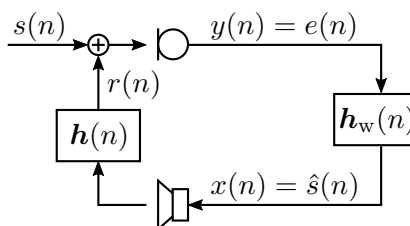


Abbildung 3.7: Wiener-Filter im Zeitbereich zur Reduktion der Rückkopplung.

vorangegangenen Abschnitte zu folgen, wird das Signal, welches den Vorwärtspfad des ICC-Systems durchläuft, mit $e(n)$ bezeichnet. Ohne den Rückkopplungskompensator entspricht dieses dem Mikrofonsignal $y(n)$. Das Wiener-Filter wird mit $h_w(n)$ bezeichnet. Der Ausgang des Wiener-Filters wird direkt am Lautsprecher ausgegeben. Er soll idealerweise nur noch das Sprachsignal $s(n)$ ohne Rückkopplungsanteile enthalten, d. h. es soll gelten

$$x(n) = \hat{s}(n) \stackrel{!}{=} s(n). \quad (3.43)$$

Äquivalent zur Bestimmung der optimalen Koeffizienten bei der Systemidentifikation (Gl. (3.17) - (3.23)) kann die Bestimmung der optimalen Koeffizienten des Wiener-Filters als Minimierung des mittleren quadratischen Fehlers zwischen $s(n)$ und $\hat{s}(n)$ formuliert werden. Um Verwechslung mit dem Fehlersignal $e(n)$ des Rückkopplungskompensators

zu vermeiden, wird für den Fehler des Wiener-Filters $e_w(n)$ der Index w eingeführt

$$\mathbb{E}\{e_w^2(n)\} = \mathbb{E}\{(\hat{s}(n) - s(n))^2\} \quad (3.44)$$

$$\mathbb{E}\{e_w^2(n)\} = \mathbb{E}\left\{\left(\mathbf{h}_w^T(n) \cdot \mathbf{e}(n) - s(n)\right)^2\right\}. \quad (3.45)$$

Bezeichnet N_{h_w} die Länge des Wiener-Filters in Abtastwerten, so gilt für den Vektor $\mathbf{e}(n)$

$$\mathbf{e}(n) = [e(n), e(n-1), \dots, e(n-N_{h_w}+1)]^T. \quad (3.46)$$

Ausmultiplizieren von Gl. (3.45) führt auf

$$\mathbb{E}\{e_w^2(n)\} = \mathbb{E}\{s^2(n)\} - 2\mathbf{r}_{se}^T(n)\mathbf{h}_w(n) + \mathbf{h}_w^T(n)\mathbf{R}_{ee}(n)\mathbf{h}_w(n). \quad (3.47)$$

Erneut sind die optimalen Koeffizienten $\mathbf{h}_{w,\text{opt}}(n)$, diejenigen, welche Gl. (3.47) minimieren. Setzt man die Ableitung von Gl. (3.47) gleich null, erhält man in diesem Fall

$$\mathbf{h}_{w,\text{opt}}(n) = \mathbf{R}_{ee}^{-1}(n) \cdot \mathbf{r}_{se}(n). \quad (3.48)$$

Gl. (3.48) wird auch als Wiener-Hopf-Gleichung bezeichnet [73]. Die Berechnung der optimalen Koeffizienten nach Gl. (3.48) erfordert die Inversion der Autokorrelationsmatrix des Fehler- bzw. in diesem Falle des Mikrofonsignals

$$\mathbf{R}_{ee}(n) = \mathbb{E}\{\mathbf{e}(n)\mathbf{e}^T(n)\}. \quad (3.49)$$

Zudem muss die lokale Sprache $s(n)$ zur Berechnung des Kreuzkorrelationsvektors

$$\mathbf{r}_{se}(n) = \mathbb{E}\{s(n)\mathbf{e}(n)\} \quad (3.50)$$

bekannt sein. Unter der Annahme, dass die lokale Sprache $s(n)$ und die Rückkopplung $r(n)$ orthogonal sind, lässt sich Gl. (3.50) auch schreiben als

$$\mathbf{r}_{se}(n) = \mathbb{E}\{s(n) \cdot (\mathbf{s}(n) + \mathbf{r}(n))\} = \underbrace{\mathbb{E}\{s(n)\mathbf{s}(n)\}}_{=\mathbf{r}_{ss}(n)} + \underbrace{\mathbb{E}\{s(n)\mathbf{r}(n)\}}_{=0} = \mathbf{r}_{ss}(n). \quad (3.51)$$

Aufgrund der Korrelation zwischen Lautsprechersignal und lokaler Sprache ist auch die Rückkopplung mit der lokalen Sprache korreliert, was bedeutet, dass $\mathbb{E}\{s(n)\mathbf{r}(n)\}$ im Fall des rückgekoppelten Systems nur näherungsweise null ist. Formt man Gl. (3.48) um, erkennt man, dass die einzelnen Zeilen der Autokorrelationsmatrix $\mathbf{R}_{ee}(n)$ jeweils mit dem Vektor der Filterkoeffizienten $\mathbf{h}_{w,\text{opt}}(n)$ gefaltet werden

$$\mathbf{R}_{ee}(n) \cdot \mathbf{h}_{w,\text{opt}}(n) = \mathbf{r}_{ss}(n). \quad (3.52)$$

Diese Faltung entspricht im Frequenzbereich einer Multiplikation. Damit kann Gl. (3.52) im diskreten Frequenzbereich angenähert werden durch

$$P_{EE}(\mu, k) \cdot H_{w,\text{opt}}(\mu, k) = P_{SS}(\mu, k), \quad (3.53)$$

wobei sich $H_{w,\text{opt}}(\mu, k)$ aus der Fouriertransformierten der Filterkoeffizienten $\mathbf{h}_{w,\text{opt}}(n)$ ergibt

$$\mathbf{h}_{w,\text{opt}}(n) \circ \bullet \mathbf{H}_{w,\text{opt}}(k) = [H_{w,\text{opt}}(0, k), H_{w,\text{opt}}(1, k), \dots, H_{w,\text{opt}}(N_{h_w} - 1, k)]^T. \quad (3.54)$$

Ebenso ergeben sich die Leistungsdichtespektren $P_{EE}(\mu, k)$ und $P_{SS}(\mu, k)$ aus den Fouriertransformationen der Autokorrelationen [32]. Das Leistungsdichtespektrum der lokalen Sprache kann ausgedrückt werden durch

$$P_{SS}(\mu, k) = P_{EE}(\mu, k) - P_{RR}(\mu, k). \quad (3.55)$$

Setzt man Gl. (3.55) in Gl. (3.53) ein und löst nach $H_{w,\text{opt}}(\mu, k)$ auf, erhält man

$$H_{w,\text{opt}}(\mu, k) = 1 - \frac{P_{RR}(\mu, k)}{P_{EE}(\mu, k)}. \quad (3.56)$$

Gl. (3.56) ist die Lösung der Wiener-Hopf-Gleichung im Frequenzbereich. Diese kann direkt im Frequenzbereich innerhalb einer Filterbank auf das Mikrofon- bzw. Fehlersignal angewendet werden, um die darin enthaltene Rückkopplung herauszufiltern. Zur Berechnung der Gleichung müssen die Leistungsdichtespektren $P_{RR}(\mu, k)$ und $P_{EE}(\mu, k)$ bekannt sein. Die entsprechenden Schätzverfahren werden in Kap. 5 vorgestellt.

3.4.2 Gewichtete *Overlap-Add*-Filterbank

Wie die Rückkopplungskompensation wird auch die Rückkopplungsreduktion aus Gründen der verringerten Rechenleistung im Frequenzbereich implementiert. Allerdings unterscheiden sich die Anforderungen, welche das Wiener-Filter an die verwendete Filterbankstruktur stellt, von denen der Rückkopplungskompensation. Die Koeffizienten des Wiener-Filters, welches bei der Rückkopplungsreduktion und beispielsweise auch bei der Geräuschreduktion zum Einsatz kommt, sind zeitvariant. Werden diese in einer *Overlap-Save*-Filterbank prozessiert, müssen zur Unterdrückung der zyklischen Faltungsanteile im Zeitbereich Nullen projiziert werden. Das bedeutet, dass das Wiener-Filter in den Zeitbereich transformiert wird, wo ein bestimmter Teil der Impulsantwort auf Null gesetzt wird. Anschließend wird die Impulsantwort zurück in den Frequenzbereich transformiert. Neben dem erhöhten Rechenaufwand durch die dafür benötigten DFTs/IDFTs wird auch noch eine zusätzliche Verzögerung eingefügt, da die Impulsantwort des Wiener-Filters zyklisch rotiert werden muss. Der Grund hierfür ist, dass das Wiener-Filter ein nullphasiges Filter ist. Nullphasig bedeutet, dass die Übertragungsfunktion der Impulsantwort rein reellwertig ist, der Imaginärteil ist null. Für die Impulsantwort bedeutet dies, dass betragsmäßig große Koeffizienten am Rand der Impulsantwort stehen, während die Impulsantwort in der Mitte abklingt. Damit die Koeffizienten am rechten Rand durch die Projektionen nicht auf null gesetzt werden, wird die Impulsantwort entweder zyklisch rotiert oder die Nullen werden in die Mitte der Impulsantwort projiziert [79]. In beiden Fällen entsteht dadurch jedoch eine zusätzliche Verzögerung, welche bei halber Überlappung der Blöcke $N/4$ beträgt.

Hinzu kommt, dass bei der *Overlap-Save*-Filterbank eine starke Änderung der Filterkoeffizienten zwischen zwei aufeinander folgenden Blöcken in Form von Kratzgeräuschen hörbar ist.

Aus den genannten Gründen werden für die Module im Vorwärtspfad des ICC-Systems häufig *Overlap-Add*-basierte Filterbänke mit einem gleitenden Fenster verwendet [14, 36]. Wie bei den *Overlap-Save*-Verfahren wird dabei das Eingangssignal ebenfalls in überlappende Blöcke der Länge N zerteilt. Zur Unterdrückung der zyklischen Faltungseffekte wird hier der Block des Eingangssignals mit einem Analysefenster

$$\mathbf{w}_a = [w_{a,0}, w_{a,1}, \dots, w_{a,N-1}]^T \quad (3.57)$$

gewichtet und anschließend mittels DFT der Ordnung N in den Frequenzbereich transformiert. Durch die Fensterung des Eingangssignals kann das resultierende Spektrum ohne Projektionen direkt mit der Übertragungsfunktion des Filters $\mathbf{H}_{w,\text{opt}}(k)$ multipliziert werden. Das verarbeitete Spektrum wird mit einer IDFT zurück in den Zeitbereich transformiert. Nach erneuter Gewichtung mit einem Synthesefenster

$$\mathbf{w}_s = [w_{s,0}, w_{s,1}, \dots, w_{s,N-1}]^T \quad (3.58)$$

werden die überlappenden Blöcke mittels *Overlap-Add* zusammengesetzt. Das Prinzip ist schematisch in Abb. 3.8 dargestellt.

Für eine perfekte Signalrekonstruktion müssen die Gewichtungen mit den Fenstern nach dem überlappenden Aufaddieren eins ergeben. Bei einem Rahmenversatz von $L = N/2$ bedeutet dies, dass gelten muss

$$w_{a,i} \cdot w_{s,i} + w_{a,i+L} \cdot w_{s,i+L} \stackrel{!}{=} 1, \quad i = 0, 1, \dots, L - 1. \quad (3.59)$$

Dies wird erfüllt, wenn sowohl für das Analysefenster als auch für das Synthesefenster z. B. das folgende Hanning-Fenster verwendet wird

$$w_{a,i} = w_{s,i} = \sqrt{\frac{L}{N} \cdot \left(1 - \cos\left(\frac{2\pi}{N} \cdot i\right)\right)} \quad i = 0, 1, \dots, N - 1. \quad (3.60)$$

3.5 Verwendete Filterbankstruktur

In Abschn. 3.3.2 wird die schnelle Faltung und die Implementierung des NLMS-Algorithmus im Frequenzbereich mit einer *Overlap-Save*-Filterbank beschrieben. Die *Overlap-Save*-Filterbank unterdrückt die Anteile der zyklischen Faltung vollständig, d. h. die schnelle Faltung entspricht exakt der aperiodischen Faltung im Zeitbereich. Für adaptive Rückkopplungskompensation bedeutet dies, dass das Filter theoretisch perfekt abgleichen kann.

In Abschn. 3.4.2 wird die für die Rückkopplungsreduktion verwendete gewichtete *Overlap-Add*-Filterbank vorgestellt. Da im Gegensatz zu einer *Overlap-Save*-Filterbank die zyklischen Faltungseffekte bei den gewichteten *Overlap-Add*-Filterbänken nicht vollständig unterdrückt werden, können adaptive Filter hier nicht perfekt abgleichen, es bleibt

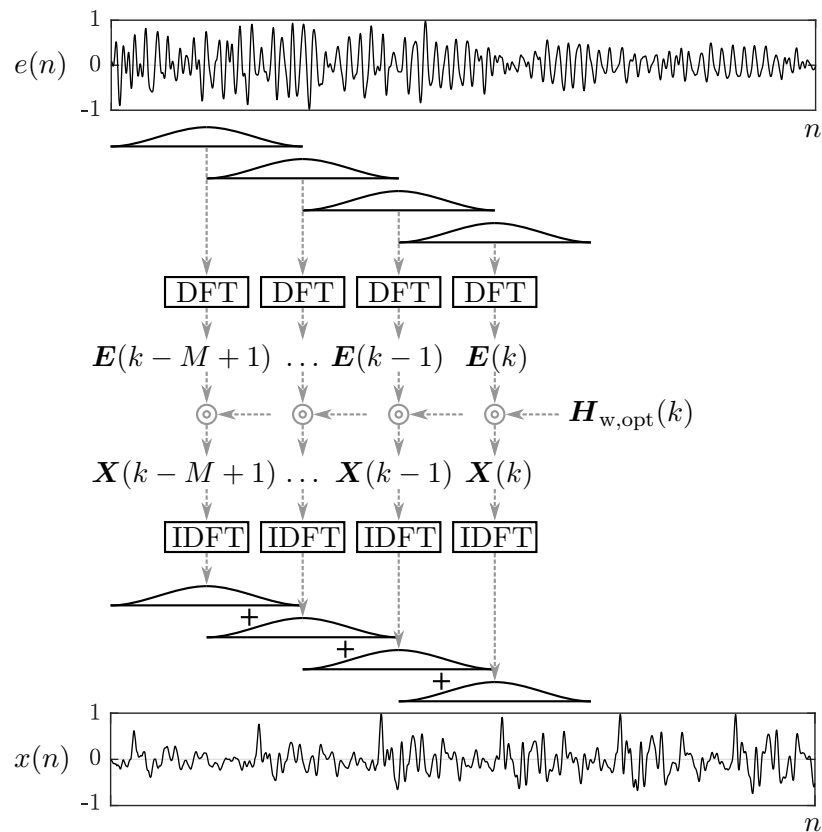


Abbildung 3.8: Blockweise Verarbeitung bei der gewichteten *Overlap-Add*-Methode. Die Blöcke überlappen sich halb.

immer ein Restfehler. Allerdings kann dieser Restfehler durch eine geeignete Auslegung der Filterbank verringert werden. Ein weiterer Nachteil ist die im Vergleich zu den *Overlap-Save*-Verfahren größere Verzögerung. Diese beträgt bei der vorgestellten *Overlap-Add*-Methode N Abtastwerte. Der Grund ist, dass gewartet werden muss, bis ein vollständiger Block verarbeitet wurde, bevor die erste Hälfte des neuen Blocks zur zweiten Hälfte des alten Blocks addiert werden kann.

Im realen Betrieb eines ICC-Systems sind neben der Rückkopplungskompensation und der Rückkopplungsreduktion zahlreiche weitere Algorithmen notwendig. Die Hauptaufgaben dieser im Folgenden *Module* genannten Algorithmen sind unter anderem:

- Reduktion der Restrückkopplungen. Immer dann, wenn die Rückkopplungskompensation schlecht abgeglichen ist, sind zusätzliche Methoden zur Reduktion der Rückkopplungsanteile notwendig (vgl. Kap. 5).
- Verbesserung des klanglichen Eindrucks durch eine automatische Entzerrung des Lautsprechersignals am Zuhörerohr (vgl. Abschn. 6.1).

- Verbesserung der Sprachverständlichkeit durch Hinzufügen von harmonischen Signalanteilen mit einem *Exciter* (vgl. Abschn. 6.2).
- Unterdrückung der Hintergrundgeräusche. Hierzu gehören beispielsweise eine Geräuschreduktion, eine geräuschabhängige Verstärkungsregelung und ein Noise Gate (vgl. z. B. [43]).
- Beeinflussung der Dynamik. Klassischerweise zählen hierzu Limiter, Kompressoren und Expander. Artverwandt sind aber auch De-Esser, welche gezielt Frequenzbereiche dämpfen, in denen Zischlaute detektiert wurden (vgl. z. B. [80, 41]).
- Verbesserung der Stabilität. Dies kann beispielsweise durch die Dekorrelation von Mikrofon- und Lautsprecher-signal oder mit adaptiven Kerbfiltern erreicht werden. Oftmals kommt für die Dekorrelation der Frequenzversatz zum Einsatz (vgl. z. B. [55, 76]).

Die Module befinden sich im Vorwärtspfad des ICC-Systems.

3.5.1 Kombination der Filterbänke

Die in den vorangegangenen Abschnitten beschriebenen Vor- und Nachteile der *Overlap-Save*- und der gewichteten *Overlap-Add*-Filterbänke sind in Hinblick auf ihre Verwendung für die Rückkopplungskompensation und die Module in Tab. 3.1 zusammengefasst.

	Vorteile	Nachteile
<i>Overlap-Save</i>	<ul style="list-style-type: none"> • Rückkopplungskompensation kann theoretisch perfekt abgleichen, keine zyklischen Faltungsanteile • Geringere Verzögerung, $L = N/2$ Abtastwerte 	<ul style="list-style-type: none"> • Erfordert Projektionen: Höherer Rechenaufwand, zusätzliche Verzögerung bei Wiener-Filtern • Hörbare Kratzgeräusche, wenn Filter sich ändert
<i>Overlap-Add gewichtet</i>	<ul style="list-style-type: none"> • Keine Projektionen notwendig, geringerer Rechenaufwand • Kratzgeräusche bei Filteränderung durch Fenster minimiert 	<ul style="list-style-type: none"> • Größere Verzögerung, N Abtastwerte • Rückkopplungskompensation gleicht nicht perfekt ab, da zyklische Faltungseffekte nicht vollständig unterdrückt werden

Tabelle 3.1: Vor-/Nachteile der Filterbankstrukturen. Die genannten Punkte sind jeweils gültig für eine Blockgröße von N und einen Rahmenversatz von $L = N/2$.

Da das Hauptaugenmerk dieser Arbeit auf der Rückkopplungskompensation liegt, soll für das adaptive Filter eine *Overlap-Save*-Filterbank verwendet werden. So wird verhindert, dass zyklische Faltungseffekte das Ergebnis beeinflussen, was eine gezieltere Untersuchung des Adaptionsprozesses erlaubt. Dennoch ist aufgrund der fehlenden Projektionen für die Module eine gewichtete *Overlap-Add*-Filterbank besser geeignet. Daher wird in dieser Arbeit für die Module eine zweite Filterbankstruktur implementiert. In der Praxis wird man einen Kompromiss eingehen und beide Teile in derselben Filterbank umsetzen.

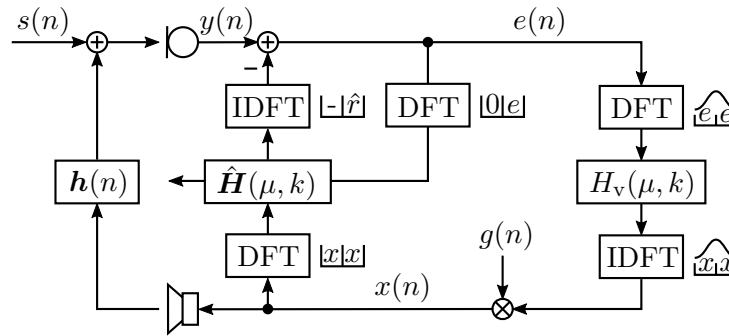


Abbildung 3.9: Blockdiagramm der in dieser Arbeit verwendeten Filterbankstruktur. Die Rückkopplungskompensation wird innerhalb einer *Overlap-Save*-Filterbank gerechnet. Im Vorwärtspfad kommt eine *Overlap-Add*-basierte Filterbank mit einem gleitenden Fenster zum Einsatz.

Die resultierende Struktur ist als Blockdiagramm in Abb. 3.9 gezeigt. Die Module befinden sich im Vorwärtspfad des Systems und sind in Abb. 3.9 in der Übertragungsfunktion $H_v(\mu, k)$ zusammengefasst. Diese befindet sich innerhalb der gewichteten *Overlap-Add*-Filterbank. Die *Overlap-Save*-Filterbank der Rückkopplungskompensation ist identisch zu der in Abb. 3.6 auf S. 34 gezeigten Struktur. Die Systemverstärkung $g(n)$ wird nach den Modulen breitbandig im Zeitbereich auf das Lautsprecher-Signal $x(n)$ angewandt. Beide Filterbänke arbeiten mit einer Blöckgröße N , bei halber Überlappung, bzw. Rahmenversatz $L = N/2$. Damit beträgt die Verzögerung im Vorwärtspfad jetzt N Abtastwerte. Man könnte auch die DFT-Länge der Module halb so lang machen wie die des Kompensators. Dann würde sich die Verzögerung im Vorwärtspfad wieder auf L Abtastwerte verkürzen. Allerdings würde dies auch in einer geringeren Frequenzauflösung der Filterbank für die Module resultieren.

Es ist außerdem in Abb. 3.9 zu erkennen, dass das Fehlersignal zweimal in den Frequenzbereich transformiert wird. Einmal wird es mittels *Overlap-Save* für den Kompensator und zusätzlich mittels gewichteter *Overlap-Add*-Methode für die Module transformiert. Letztere DFT kann eingespart werden, indem das Spektrum des vorangegangenen Blocks der *Overlap-Save*-Methode mit dem Faktor $e^{-j\frac{2\pi\mu}{N}\cdot L}$ multipliziert wird. Im Zeitbereich entspricht dies einer zyklischen Rotation um L Werte. Anschließend addiert man das rotierte Spektrum zum aktuellen und faltet das Resultat mit der DFT des verwendeten modifizierten Hanning-Fensters w_a . Man erhält so das Fehlersignal im Frequenz-

bereich, wie es für die gewichtete *Overlap-Add*-Methode benötigt wird. Mathematisch ausgedrückt bedeutet dies

$$\mathbf{E}_{\text{OA}}(k) = \left(\mathbf{E}_{\text{OS}}(k) + \mathbf{E}_{\text{OS}}(k-1) \circ \left[W_N^{0:L}, W_N^{1:L}, \dots, W_N^{N-1:L} \right]^T \right) * \text{DFT}\{\mathbf{w}_a\} \quad (3.61)$$

wobei W_N der komplexe Drehfaktor

$$W_N = e^{-j\frac{2\pi}{N}} \quad (3.62)$$

ist und $*$ eine Faltung bezeichnet. Der Index OS kennzeichnet hier ein mittels *Overlap-Save* in den Frequenzbereich transformiertes Signal, analog kennzeichnet OA die gewichtete *Overlap-Add*-Filterbank. Der besseren Lesbarkeit wegen werden diese Indizes im weiteren Verlauf dieser Arbeit nicht weiter verwendet. Es sei daher an dieser Stelle nochmals ausdrücklich hervorgehoben, dass sämtliche für die adaptive Rückkopplungskompensation benötigten Signale mittels *Overlap-Save* in den Frequenzbereich transformiert werden, während die Module in einer gewichteten *Overlap-Add*-Filterbank verarbeitet werden.

3.5.2 Parametrierung

In Abschn. 2.2 wurden Totzeit und Nachhallzeit verschiedener Impulsantworten im Fahrzeug bestimmt. In einem ICC-System, welches die Sprache von vorne nach hinten verstärken soll, muss die Rückkopplungskompensation die Impulsantwort von den hinteren Lautsprechern zu den vorderen Mikrofonen schätzen. Nach Abb. 2.10b auf S. 14 beträgt die Totzeit dieses Pfades beim vollbesetzten Fahrzeug $T_D = 5,6$ ms, die Nachhallzeit ist $T_{60} = 80,6$ ms. Die Gesamtdauer, ab der die Energie der Impulsantwort unter -60 dB abgefallen ist, beträgt also in diesem Beispiel 86,2 ms.

Die Konvergenz eines adaptiven Filters hängt von der Länge des Filters ab. Je länger das Filter ist, desto mehr Koeffizienten müssen geschätzt werden. Das führt zu einer langsameren Konvergenz und zum anderen zu einer höheren Rechenlast als bei einer kurzen Filterlänge. Auf der anderen Seite kann bei einer größeren Anzahl korrekt geschätzter Koeffizienten die Impulsantwort präziser nachgebildet werden [29]. Vor allem bei der Echokompensation spielt dies eine große Rolle, da die Eigenschaften des Raumes, in dem das System betrieben wird, nicht bekannt sind und sich z. B. bei mobilen Geräten stark ändern können. Eine Lösung zur Bestimmung der optimalen Filterlänge ist das in [59] vorgestellte Verfahren. Bei diesem wird die Filterlänge eines Echokompensators zur Laufzeit adaptiv angepasst.

Dem gegenüber kann in einem bekannten Fahrzeug die Länge der Impulsantwort vorab gemessen werden und ändert sich nur in geringem Rahmen. Daher soll in der vorliegenden Arbeit eine feste Länge verwendet werden. In der Praxis hat sich als guter Kompromiss herausgestellt, die Impulsantwort bis zu dem Zeitpunkt nachzubilden, ab dem die EDC unter -30 dB abgefallen ist. Setzt man einen logarithmisch linearen Verlauf der EDC voraus, entspricht dieser T_{30} genannte Zeitpunkt der halben Nachhallzeit T_{60} .

Addiert man also im oben beschriebenen Fall des vollbesetzten Fahrzeugs die T_{30} zur Totzeit T_D erhält man die Länge der zu schätzenden Impulsantwort. Bei einer Abtastfrequenz von $f_s = 44,1$ kHz hat der relevante Teil der zu schätzenden Impulsantwort damit

eine Länge von

$$\begin{aligned} N_h &= (T_D + T_{30}) \cdot f_s \\ &= (5,6 \text{ ms} + 40,3 \text{ ms}) \cdot 44,1 \text{ kHz} \approx 2024 \end{aligned} \quad (3.63)$$

Abtastwerten. Aus rechentechnischen Gründen wird eine Potenz von zwei bevorzugt, sodass im Folgenden mit einer Filterlänge von $N_h = 2048$ Koeffizienten gearbeitet wird.

Damit die Verzögerung im Vorwärtspfad kleiner 15 ms bleibt, wird eine Blockgröße von $N = 512$ Abtastwerten gewählt. Bei der genannten Abtastfrequenz entspricht dies einer Verzögerung von ca. 11,6 ms. Da sich die Blöcke halb überlappen, beträgt der Rahmenversatz folglich $L = 256$ Abtastwerte. Für das adaptive Filter bedeutet dies, dass

$$M = \frac{N_h}{L} = \frac{2048}{256} = 8 \quad (3.64)$$

Teilfilter benötigt werden, um den relevanten Teil der Impulsantwort abzudecken.

Die hier bestimmten Parameter sind in Tab. 3.2 zusammengefasst und werden, wenn nicht anders angegeben, für die in den nachfolgenden Kapiteln durchgeführten Simulationen verwendet.

Parameter	Wert
Abtastfrequenz	$f_s = 44,1 \text{ kHz}$
Länge DFT	$N = 512$
Rahmenversatz	$L = N/2 = 256$
Verzögerung im Vorwärtspfad	$D = N = 512$ (ca. 11,6 ms)
Länge reale (simulierte) Impulsantwort	$N_h = 2048$
Anzahl Teilfilter	$M = 8$
Länge geschätzte Impulsantwort	$N_{\hat{h}} = M \cdot L = 2048$

Tabelle 3.2: In dieser Arbeit verwendete Simulationsparameter.

3.6 Zusammenfassung

In diesem Kapitel wurde zunächst das Stabilitätskriterium für rückgekoppelte elektroakustische Systeme hergeleitet. Nach einem Abschnitt über den Stand der Technik wurde anschließend gezeigt, wie die Rückkopplung mit einem adaptiven Filter prinzipiell kompensiert werden kann. Im nachfolgenden Abschnitt über das Prinzip der adaptiven Rückkopplungskompensation wurde erläutert, weshalb eine vollständige Kompensation der Rückkopplung aufgrund der starken Korrelation der Signale ohne zusätzliche Maßnahmen nicht realisierbar ist. Der Abschnitt stellt die Grundlage für Kap. 4 dar, in welchem eine Schrittweitensteuerung zur Verbesserung der Konvergenz eines Rückkopplungskompensators beschrieben ist. Als Alternative bzw. Ergänzung zur Rückkopplungskompensation wurde außerdem das Prinzip der Rückkopplungsreduktion mittels Wiener-Filter

beschrieben. Das in Kap. 5 erläuterte Verfahren basiert auf dieser Herleitung. Sowohl die Rückkopplungskompensation als auch die Rückkopplungsreduktion können zur Verringerung der Rechenlast im Frequenzbereich implementiert werden. Beide Verfahren haben jeweils unterschiedliche Anforderungen an die dafür benötigten Filterbänke. Die Vor- und Nachteile der jeweiligen Filterbänke wurden aufgezeigt und abschließend die im weiteren Verlauf der Arbeit verwendete Filterbankstruktur vorgestellt.

Kapitel 4

Schrittweitensteuerung für die adaptive Rückkopplungskompensation

Dieses Kapitel stellt den Kern der vorliegenden Arbeit dar. Es beschreibt das entwickelte Verfahren zur Steuerung der Schrittweite bei der akustischen Rückkopplungskompensation. Zunächst wird dafür die theoretisch optimale Schrittweite für adaptive Filter hergeleitet. Anschließend wird das Prinzip der Nachhall-basierten Schrittweitensteuerung vorgestellt. Es zeigt sich, dass die Nachhall-basierte Schrittweite als Approximation der optimalen Schrittweite interpretiert werden kann, wenn die Verstärkung im Vorwärtspfad des Systems mit dem inversen Systemabstand gesteuert wird. Im weiteren Verlauf des Kapitels werden die dafür benötigten Kontrollmechanismen erläutert. Dazu zählen insbesondere Verfahren zur Schätzung des Systemabstands und damit zusammenhängend Mechanismen zur Detektion von Raumänderungen. Die Simulationsergebnisse werden der Übersichtlichkeit halber am Ende des Kapitels präsentiert. Teile dieses Kapitels wurden bereits in [6, 7, 8] veröffentlicht.

4.1 Optimale Schrittweite

In Abschn. 3.3.2 wird die Implementierung des NLMS-Algorithmus im Frequenzbereich erläutert. Gl. (3.40) beschreibt die rekursive Aktualisierung des adaptiven Filters. Wird das adaptive Filter innerhalb einer *Overlap-Save*-Filterbank umgesetzt, sind Projektionen beim Gradient notwendig, um zyklische Faltungseffekte zu unterdrücken. Im Folgenden wird die optimale Schrittweite für das adaptive Filter hergeleitet. Um die Projektionen bei dieser Herleitung nicht betrachten zu müssen, wird davon ausgegangen, dass die Signale bereits ohne zyklische Faltungseffekte im Frequenzbereich vorliegen. Aus diesem Grund sind die Signale in Abb. 4.1 im Frequenzbereich dargestellt, die Filterbänke werden nicht gezeigt.

Es folgt damit für die Aktualisierung des adaptiven Filters ohne Projektionen

$$\hat{\mathbf{H}}(\mu, k+1) = \hat{\mathbf{H}}(\mu, k) + \alpha(\mu, k) \cdot \frac{E(\mu, k) \cdot \mathbf{X}^*(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}. \quad (4.1)$$

Wie bereits in Kap. 3 erläutert wurde, beschreibt $\alpha(\mu, k)$ die Schrittweite, mit welcher man sich entlang der Kostenfunktion dem Minimum nähert. Je kleiner die Schrittweite ist, desto präziser wird das Minimum gefunden, allerdings auf Kosten einer langsameren Konvergenzgeschwindigkeit. Umgekehrt bedeutet eine große Schrittweite, dass das Filter zwar schneller konvergieren kann, dafür jedoch schlechter abgleicht. Optimal ist daher

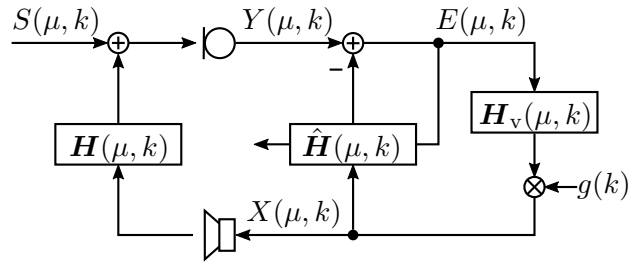


Abbildung 4.1: Blockdiagramm der adaptiven Rückkopplungskompensation. Zur vereinfachten Darstellung sind sämtliche Signale im Frequenzbereich gezeigt.

eine geregelte Schrittweite. Diese muss am Anfang des Adaptionsprozesses groß sein und kleiner werden, je besser das System abgeglichen ist.

4.1.1 Herleitung der optimalen Schrittweite

Das Optimierungskriterium zur Bestimmung der optimalen Schrittweite ist der Systemabstand. Dieser beschreibt den Abstand zwischen realem System und adaptivem Filter. Das reale System wird in diesem Fall durch die Teilbandimpulsantwort $\mathbf{H}(\mu, k)$ beschrieben, welche aus der blockweisen DFT der realen Impulsantwort $\mathbf{h}(n)$ bestimmt werden kann. Der Systemabstand ist die quadratische Norm der Differenz zwischen den Koeffizienten von realer und geschätzter Teilbandimpulsantwort

$$\|\mathbf{H}_{\Delta}(\mu, k)\|^2 = \|\mathbf{H}(\mu, k) - \hat{\mathbf{H}}(\mu, k)\|^2. \quad (4.2)$$

Läuft $\hat{\mathbf{H}}(\mu, k) \rightarrow \mathbf{H}(\mu, k)$ wird der Systemabstand minimiert. Die optimale Schrittweite ist diejenige, mit welcher dies am schnellsten gelingt.

Nachfolgend wird die optimale Schrittweite $\alpha_{\text{opt}}(\mu, k)$ hergeleitet. Dabei werden der Übersicht halber nur die wichtigsten Schritte aufgezeigt. Eine detaillierte Herleitung mit allen Zwischenschritten ist in Anhang A zu finden.

Zunächst setzt man Gl. (4.1) in Gl. (4.2) ein und erhält

$$\|\mathbf{H}_{\Delta}(\mu, k+1)\|^2 = \|\mathbf{H}_{\Delta}(\mu, k) - \alpha(\mu, k) \cdot \frac{E(\mu, k) \cdot \mathbf{X}^*(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}\|^2. \quad (4.3)$$

Der Systemabstand ist also eine quadratische Funktion der Schrittweite, an deren Minimum die optimale Schrittweite liegt. Dieses Minimum wird berechnet, indem der Erwartungswert des Systemabstands zum Zeitpunkt $k+1$ nach der Schrittweite abgeleitet und anschließend null gesetzt wird

$$\frac{\partial \mathbb{E}\{\|\mathbf{H}_{\Delta}(\mu, k+1)\|^2\}}{\partial \alpha(\mu, k)} \stackrel{!}{=} 0. \quad (4.4)$$

Auflösen von Gl. (4.4) nach der Schrittweite führt dann auf

$$\alpha_{\text{opt}}(\mu, k) = \frac{\mathbb{E}\{|E_u(\mu, k)|^2\}}{\mathbb{E}\{|E(\mu, k)|^2\}}. \quad (4.5)$$

Dies ist die bekannte optimale Schrittweite, wie sie z.B. auch in [29, 31] hergeleitet worden ist. Zur Berechnung dieser optimalen Schrittweite werden sowohl der gestörte Fehler $E(\mu, k)$ als auch der ungestörte Fehler $E_u(\mu, k)$ benötigt. Der gestörte Fehler $E(\mu, k)$ ist direkt messbar. Er setzt sich nach Abb. 4.1 zusammen aus

$$E(\mu, k) = S(\mu, k) + \mathbf{X}^T(\mu, k)\mathbf{H}(\mu, k) - \mathbf{X}^T(\mu, k)\hat{\mathbf{H}}(\mu, k). \quad (4.6)$$

Im Gegensatz dazu ist der ungestörte Fehler $E_u(\mu, k)$ unbekannt und muss geschätzt werden. Er ist immer dann messbar, wenn keine lokale Sprache am Mikrofon anliegt, d. h. wenn $S(\mu, k) = 0$. Betrachtet man erneut Abb. 4.1, so erkennt man, dass für den ungestörten Fehler im Fall $S(\mu, k) = 0$ gilt

$$E_u(\mu, k) = \mathbf{X}^T(\mu, k)\mathbf{H}(\mu, k) - \mathbf{X}^T(\mu, k)\hat{\mathbf{H}}(\mu, k) \quad (4.7)$$

$$= \mathbf{X}^T(\mu, k) \cdot (\mathbf{H}(\mu, k) - \hat{\mathbf{H}}(\mu, k)) \quad (4.8)$$

$$= \mathbf{X}^T(\mu, k) \cdot \mathbf{H}_\Delta(\mu, k). \quad (4.9)$$

Setzt man den ungestörten Fehler in die optimale Schrittweite nach Gl. (4.5) ein, ergibt sich

$$\alpha_{\text{opt}}(\mu, k) = \frac{\mathbb{E}\{|\mathbf{X}^T(\mu, k) \cdot \mathbf{H}_\Delta(\mu, k)|^2\}}{\mathbb{E}\{|E(\mu, k)|^2\}}. \quad (4.10)$$

4.1.2 Näherungen und zu schätzende Größen

In [31] wird an dieser Stelle die Annahme getroffen, dass das Lautsprechersignal $|X(\mu, k)|^2$ und der Systemabstand $\|\mathbf{H}_\Delta(\mu, k)\|^2$ unkorreliert sind, wodurch der Zähler in Gl. (4.10) wie folgt umgeformt werden kann

$$\mathbb{E}\{|\mathbf{X}^T(\mu, k) \cdot \mathbf{H}_\Delta(\mu, k)|^2\} \approx \mathbb{E}\{|X(\mu, k)|^2\} \cdot \mathbb{E}\{\|\mathbf{H}_\Delta(\mu, k)\|^2\}, \quad (4.11)$$

woraus für die optimale Schrittweite folgt

$$\alpha_{\text{opt}}(\mu, k) \approx \frac{\mathbb{E}\{|X(\mu, k)|^2\}}{\mathbb{E}\{|E(\mu, k)|^2\}} \cdot \mathbb{E}\{\|\mathbf{H}_\Delta(\mu, k)\|^2\}. \quad (4.12)$$

Es ist zu beachten, dass im Fall der geschlossenen Schleife bei der Rückkopplungskompensation diese Näherung zunächst nicht gültig ist. Es wird sich allerdings im weiteren Verlauf dieses Kapitels zeigen, dass es mit speziellen Kontrollmechanismen möglich ist, mit Gl. (4.12) die Schrittweite auch in einer geschlossenen Schleife zu steuern.

Dass die Abhängigkeit der Schrittweite $\alpha_{\text{opt}}(\mu, k)$ vom Systemabstand $\|\mathbf{H}_\Delta(\mu, k)\|^2$ in einem optimalen Konvergenzverhalten resultiert, lässt sich auch anschaulich erklären. Am Beginn des Adaptionsprozesses ist das Filter schlecht abgeglichen, d. h. der Systemabstand ist groß. Damit wird auch die Schrittweite groß und das Filter konvergiert zwar ungenau, dafür aber schnell. Je besser das Filter abgeglichen ist, desto kleiner wird der Systemabstand und damit auch die Schrittweite, wodurch das Filter immer präziser gegen die optimale Lösung konvergieren kann [29]. Ändert sich der Raum, d. h. die

zu schätzende Impulsantwort, z. B. durch schnelle Bewegungen der Passagiere, wird der Systemabstand schlagartig groß. Dadurch wird auch die Schrittweite groß und das Filter kann schnell gegen die neue Impulsantwort konvergieren.

Da sowohl der Systemabstand $\|\mathbf{H}_\Delta(\mu, k)\|^2$ als auch die Erwartungswerte $E\{\cdot\}$ unbekannt sind, müssen diese geschätzt werden. Ein Verfahren zur Schätzung des Systemabstands wird in Abschn. 4.4 vorgestellt. Die Erwartungswerte, bzw. die mittleren Leistungen, können angenähert werden durch Glättung der Momentanleistung mit einem IIR-Filter erster Ordnung

$$E\{|X(\mu, k)|^2\} \approx \overline{|X(\mu, k)|^2} = \beta \cdot \overline{|X(\mu, k-1)|^2} + (1 - \beta) \cdot |X(\mu, k)|^2, \quad (4.13)$$

wobei β die Glättungskonstante bezeichnet, welche im Bereich $0 \dots 1$ liegt [29]. Die gleiche Näherung kann auch für $E(\mu, k)$ und $\mathbf{H}_\Delta(\mu, k)$ angewendet werden, sodass Gl. (4.12) schließlich angenähert wird durch

$$\alpha_{\text{opt}}(\mu, k) \approx \frac{\overline{|X(\mu, k)|^2}}{\overline{|E(\mu, k)|^2}} \cdot \overline{\|\mathbf{H}_\Delta(\mu, k)\|^2}. \quad (4.14)$$

4.2 Nachhall-basierte Schrittweitensteuerung

In diesem Abschnitt wird ein Verfahren zur Steuerung der Schrittweite des adaptiven Filters für die Rückkopplungskompensation vorgestellt. Die grundlegende Idee dieses Verfahrens hat Ähnlichkeit mit den Mechanismen zur Steuerung der Schrittweite bei der akustischen Echokompensation. Bei der Echokompensation muss zu den Zeitpunkten, während derer beide Sprecher aktiv sind (Gegensprechen), die Schrittweite verringert werden, um eine Divergenz des Filters zu verhindern. Sobald nur noch der ferne Teilnehmer aktiv ist, wird die Schrittweite erhöht, da dann lediglich das Anregungssignal am Echokompensator anliegt und dieser somit zur korrekten Lösung konvergieren kann. Verschiedene Verfahren zur Steuerung der Schrittweite und insbesondere zur Detektion von Gegensprechen bei der Echokompensation sind z. B. in [46] beschrieben.

Auch das adaptive Filter eines Rückkopplungskompensators kann ungestört konvergieren, wenn ausschließlich ein Signal am Lautsprecher anliegt, aber keine lokale Sprachaktivität. Da jedoch das Lautsprechersignal idealerweise der lokalen Sprache entsprechen soll, befindet sich ein rückgekoppeltes System zunächst einmal permanent im Zustand des Gegensprechens (vgl. Abschn. 3.3.1). Bei näherer Betrachtung stellt man allerdings fest, dass es auch in einem rückgekoppelten System kurze Zeitpunkte gibt, während derer der lokale Sprecher inaktiv ist, jedoch am Lautsprecher noch ein Signal anliegt. Diese Zeitpunkte treten immer unmittelbar am Ende lokaler Sprechaktivität auf und haben als Ursache den Nachhall des Systems.

4.2.1 Nachhall

Klassischerweise wird mit Nachhall die Energie bezeichnet, welche unmittelbar nach dem Abschalten einer Schallquelle im Raum verbleibt und durch Reflexionen an den Wänden und Gegenständen im Inneren des Raumes exponentiell abklingt [37]. Der Nachhall

wird dabei beeinflusst von der Größe und der Beschaffenheit des Raumes. So klingt die Energie nach dem Abschalten eines Schallereignisses in großen Hallen mit vielen harten, schallreflektierenden Flächen, wie z. B. Kirchen, nur sehr langsam ab, was in einem deutlich hörbaren, langen Nachhall resultiert. In kleineren, gedämpften Räumen, zu denen auch Fahrzeuginnenräume gehören, klingt ein Schallereignis deutlich schneller ab. Die akustischen Eigenschaften eines Fahrzeuginnenraums wurden bereits in Abschn. 2.2 analysiert. Die EDCs beschreiben dabei das am Mikrofon gemessene Abklingverhalten des Raumes. Unter anderem wurde in dem Abschnitt gezeigt, dass die Nachhallzeit umso kürzer wird, je mehr Passagiere sich im Fahrzeug befinden, da weiche Gegenstände, wie z. B. Stoffe, den Schall absorbieren.

Wird nun in einem Raum eine elektro-akustische Schleife bestehend aus Lautsprechern und Mikrofonen betrieben, wird der Nachhall hiervon zusätzlich beeinflusst. Grund sind die Verzögerung und die Verstärkung im Vorwärtspfad des Systems sowie die Laufzeit zwischen Lautsprecher und Mikrofon. Je länger die Summe aus Verzögerung und Laufzeit wird, desto länger klingt auch das System nach. Dies lässt sich folgendermaßen veranschaulichen: Ein vom Mikrofon aufgenommenes Schallereignis wird im Vorwärtspfad verzögert und anschließend über den Lautsprecher wiedergegeben. Auf dem direkten Weg vom Lautsprecher zum Mikrofon erfährt das Signal eine weitere Verzögerung, welche abhängig ist von der Schallgeschwindigkeit und der zurückgelegten Strecke. Ohne weitere Maßnahmen zur Kompensation der Rückkopplung wird das verzögerte Signal nun abermals vom Mikrofon aufgenommen und durchläuft so mehrmals die Schleife. Vom Lautsprecher zum Mikrofon wird das Signal gedämpft, während es im Vorwärtspfad verstärkt wird. Je größer die Verstärkung ist, desto lauter sind die rückgekoppelten Signalanteile am Mikrofon, was den Nachhall ebenfalls verlängert.

Es muss also unterschieden werden, durch welchen Effekt der Nachhall verursacht wird. Während sich der Nachhall, der von den akustischen Eigenschaften des Raumes abhängt, mittels Signalverarbeitung nicht beeinflussen lässt, kann der durch die Schleife zusätzlich hervorgerufene Nachhall ausgenutzt werden. Wenn im weiteren Verlauf dieses Kapitels von Nachhall die Rede ist, sind daher damit diejenigen Schallanteile gemeint, welche unmittelbar nach einem Schallereignis im Inneren des Fahrzeugs am Lautsprecher ausgegeben werden.

Dieser am Lautsprecher auftretende Nachhall ist in Abb. 4.2 visualisiert. In der Abbildung sind zwei Spektrogramme gezeigt. Das obere Spektrogramm zeigt ein lokales Schallereignis im Inneren des Fahrzeugs. Im unteren Spektrogramm ist das resultierende Lautsprechersignal dargestellt. Für das lokale Signal wird hier ein weißes Rauschen der Länge 0,5 s verwendet. Dieses weiße Rauschen sorgt dafür, dass alle Frequenzen im System gleichmäßig angeregt werden. Der für die Simulation verwendete Rückkopplungspfad ist der Pfad zwischen den hinteren Lautsprechern und dem vorderen linken Mikrofon (vgl. Abb. 2.8). Die Systemverstärkung wurde so eingestellt, dass das System ca. 5 dB unterhalb der Stabilitätsgrenze betrieben wird. Man erkennt deutlich, dass nachdem die lokale Sprache stoppt, das Lautsprechersignal noch für über 0,5 s nachhallt, obwohl für die offene Schleife die Nachhallzeit für diesen Pfad zu $T_{60} = 120,2$ ms bestimmt worden ist. D. h. die geschlossene Schleife sorgt wie beschrieben dafür, dass die Nachhallzeit des Raumes künstlich verlängert wird. Es ist weiter zu sehen, dass das Abklingverhalten

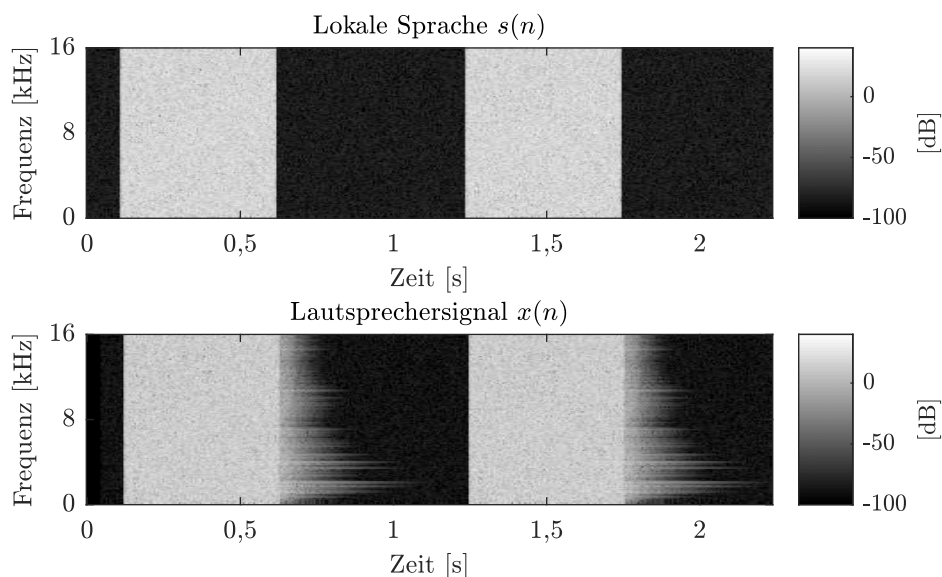


Abbildung 4.2: Der Nachhall in einer geschlossenen elektro-akustischen Schleife. Als lokales Schallereignis wird weißes Rauschen verwendet. Nachdem das Rauschen stoppt, klingt das Lautsprechersignal langsam ab.

über die Frequenz variiert, was mit dem Übertragungsverhalten des Systems zusammenhängt (vgl. Abb. 2.7). Frequenzen, die im Raumfrequenzgang stärker gedämpft werden, klingen dementsprechend weniger lang nach.

4.2.2 Detektion des Nachhalls

Im vorangegangenen Abschnitt wurde gezeigt, dass während des Nachhalls keine lokale Sprachaktivität vorliegt ($s(n) = 0$), am Lautsprecher jedoch noch das Anregungssignal $x(n)$ anliegt. Das bedeutet, dass in Gl. (3.23) der Kreuzkorrelationsvektor zwischen Anregungssignal und Lautsprechersignal null wird

$$\mathbf{r}_{sx}(n) = \mathbf{0}_{N_h}. \quad (4.15)$$

Gelingt es also, das Filter während des Nachhalls konvergieren zu lassen und während lokaler Sprachaktivität den Adaptionsprozess anzuhalten, kann das Filter ohne zusätzliche Dekorrelationsmethoden gegen die korrekte Lösung

$$\hat{\mathbf{H}}(\mu, k) \rightarrow \mathbf{H}(\mu, k) \quad (4.16)$$

konvergieren. Der Adaptionsprozess wird in Gl. (4.1) über die Schrittweite $\alpha(\mu, k)$ eingestellt. Ziel ist es nun, die Schrittweite so zu steuern, dass diese im Nachhall erhöht wird und am Beginn bzw. während lokaler Sprachaktivität gestoppt bzw. verringert wird. Es wird also ein Mechanismus benötigt, mit welchem der Nachhall detektiert werden kann. Hierfür werden im Folgenden zwei Eigenschaften der verwendeten Filterbankstruktur ausgenutzt:

1. Das Lautsprechersignal $x(n)$ entspricht dem verstärkten und aufgrund der Filterbank im Vorwärtspfad um N Abtastwerte verzögerten Fehlersignal $e(n)$. Im ersten Schritt sollen die Module im Vorwärtspfad nicht betrachtet werden. D. h. in Abb. 3.9 auf S. 42 gilt $H_v(\mu, k) = 1$ und damit

$$x(n) = g(n) \cdot e(n - N). \quad (4.17)$$

Diese Eigenschaft sorgt dafür, dass das Fehlersignal Änderungen im Vorwärtspfad *früher* folgt, als das Lautsprechersignal.

2. Die linke Hälfte des Fehlerblocks wird bei der *Overlap-Save*-Struktur zu null gesetzt, bevor der Block für die Rückkopplungskompensation mittels DFT in den Frequenzbereich transformiert wird (vgl. G. (3.38)). Das bedeutet, dass das Rechteckfenster über dem Fehlersignal halb so lang ist wie das Rechteckfenster über dem Lautsprechersignal. Dies hat zur Folge, dass die Teilbandsignale $E(\mu, k)$ Änderungen im Vorwärtspfad *schneller* folgen können als die Teilbandsignale des Lautsprechersignals $X(\mu, k)$.

Die beiden beschriebenen Eigenschaften sind schematisch in Abb. 4.3 gezeigt. Im obe-

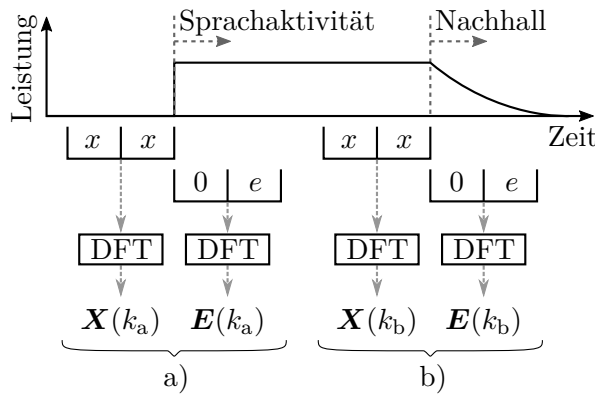


Abbildung 4.3: Das Prinzip der Nachhalldetektion. Der Block des Fehlersignals eilt dem des Lautsprechersignals voraus und wird mit einem kürzeren Fenster gewichtet. Der Zeitpunkt a) markiert den Beginn von Sprachaktivität, b) den Nachhall.

ren Teil des Bildes ist die Leistung im Vorwärtspfad zu sehen. Während Sprachaktivität nimmt diese zu und klingt am Ende der Sprachaktivität als Nachhall ab. Die Zeitbereichsblöcke, welche mittels DFT in den Frequenzbereich transformiert werden, sind unter der Kurve gezeigt. Setzt man die geglätteten Leistungen von Fehler- und Lautsprechersignal zum Zeitpunkt k ins Verhältnis, erhält man

$$\alpha_{\text{hall}}(\mu, k) = \frac{\overline{|X(\mu, k)|^2}}{g^2(k) \cdot \overline{|E(\mu, k)|^2}}, \quad (4.18)$$

wobei $\overline{(\cdot)}$ eine IIR-Glättung gemäß Gl. (4.13) bezeichnet. In Gl. (4.17) wird das Fehlersignal mit $g(n)$ multipliziert, um das Lautsprechersignal zu erhalten. Aus diesem Grund

muss die mittlere Fehlerleistung mit der quadratischen Verstärkung $g^2(k)$ multipliziert werden, damit die Leistungen von $X(\mu, k)$ und $E(\mu, k)$ im Vorwärtspfad vergleichbar bleiben. In Abb. 4.3 sind zwei Zeitpunkte a) und b) markiert, welche im Folgenden genauer betrachtet werden sollen.

Der Zeitpunkt a) markiert den Beginn von Sprachaktivität. Die mittlere Fehlerleistung beinhaltet bereits Sprachanteile, während im Lautsprecher noch keine Sprache enthalten ist. Für die Nachhalldetektion nach Gl. (4.18) gilt in diesem Fall

$$\alpha_{\text{hall}}(\mu, k_a) = \frac{\overline{|X(\mu, k_a)|^2}}{g^2(k_a) \cdot \overline{|E(\mu, k_a)|^2}} < 1. \quad (4.19)$$

Mit b) ist das Ende von Sprachaktivität gekennzeichnet. Das Fehlersignal liegt bereits im Nachhall, während am Lautsprecher noch Sprache anliegt. Es folgt für die Nachhalldetektion

$$\alpha_{\text{hall}}(\mu, k_b) = \frac{\overline{|X(\mu, k_b)|^2}}{g^2(k_b) \cdot \overline{|E(\mu, k_b)|^2}} > 1. \quad (4.20)$$

Nicht markiert sind die Zeitpunkte k_c , während denen konstant lokale Sprache anliegt, bzw. der Sprecher über einen längeren Zeitraum nicht aktiv ist. Während dieser Perioden ist die Leistung im Vorwärtspfad näherungsweise konstant, für die Nachhalldetektion gilt dementsprechend

$$\alpha_{\text{hall}}(\mu, k_c) = \frac{\overline{|X(\mu, k_c)|^2}}{g^2(k_c) \cdot \overline{|E(\mu, k_c)|^2}} \approx 1. \quad (4.21)$$

Der zeitliche Verlauf der Größen für die Nachhalldetektion ist in Abb. 4.4 dargestellt. Zur Simulation der Leistung eines Sprachsignals im Vorwärtspfad wurde ein Rechteckimpuls verwendet, dessen rechte Flanke exponentiell abklingt. Dies soll den Nachhall simulieren. Die aufgrund der Blockverarbeitung diskretisierten Leistungen $\overline{|X(\mu, k)|^2}$ und $\overline{|E(\mu, k)|^2}$ sind im mittleren Bild der Abbildung gezeigt, wobei die Verstärkung $g(k)$ in diesem Fall 0 dB beträgt. Im unteren Bild ist schließlich der resultierende Verlauf der Nachhalldetektion $\alpha_{\text{hall}}(\mu, k)$ dargestellt. Diese weist wie erwartet eine deutliche Kerbe am Beginn sowie eine Überhöhung am Ende der Sprachaktivität auf.

Man erkennt also, dass mit Gl. (4.18) der Nachhall detektiert werden kann. Näherungsweise immer dann, wenn $\alpha_{\text{hall}}(\mu, k)$ den Wert eins übersteigt, befindet sich das System im Nachhall. Um mit dieser Nachhalldetektion die Schrittweite $\alpha(\mu, k)$ der Filteradaptation nach Gl. (4.1) zu steuern, kann $\alpha_{\text{hall}}(\mu, k)$ auf Minimal- und Maximalwerte α_{min} und α_{max} begrenzt und anschließend mit einem konstanten Wert α_{fix} multipliziert werden

$$\alpha(\mu, k) = \alpha_{\text{fix}} \cdot \begin{cases} \alpha_{\text{min}} & \text{wenn } \alpha_{\text{hall}}(\mu, k) < \alpha_{\text{min}}, \\ \alpha_{\text{max}} & \text{wenn } \alpha_{\text{hall}}(\mu, k) > \alpha_{\text{max}}, \\ \alpha_{\text{hall}}(\mu, k) & \text{sonst.} \end{cases} \quad (4.22)$$

Die Minimal- und Maximalwerte sorgen dafür, dass die Schrittweite innerhalb eines plausiblen Wertebereiches gehalten wird. So kann mit α_{min} verhindert werden, dass die Schrittweite zu klein wird. Allerdings hat sich gezeigt, dass dies nicht notwendig ist,

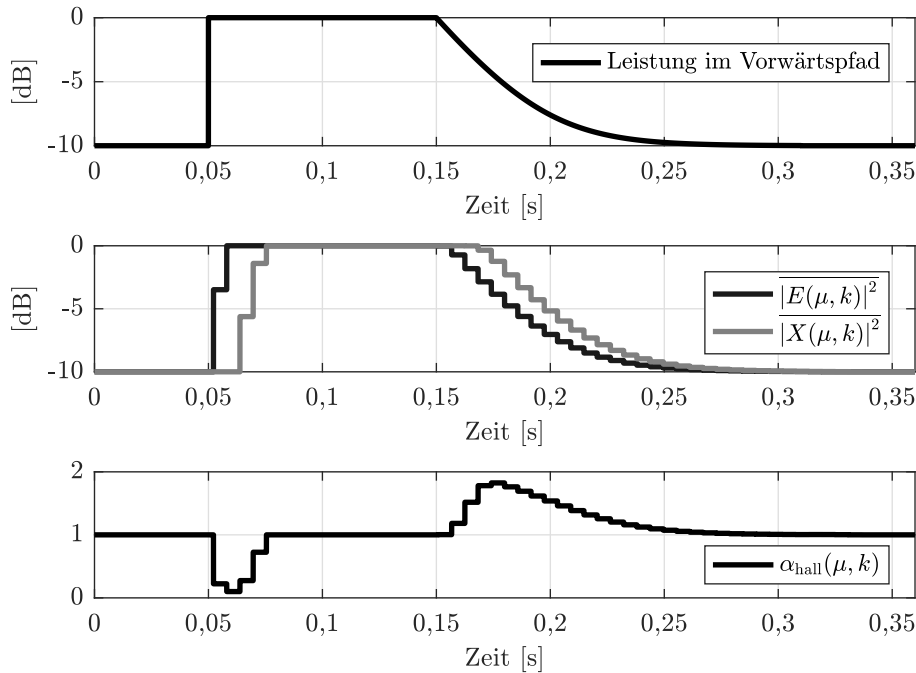


Abbildung 4.4: Verläufe der Größen für die Nachhalldetektion. Die Leistung im Vorwärtspfad wird durch einen Rechteckimpuls simuliert, dessen rechte Flanke exponentiell abklingt.

weshalb α_{\min} im Folgenden auf den Wert null gesetzt wird. Für den Maximalwert hat sich in Simulationen herausgestellt, dass Werte im Bereich $\alpha_{\max} = 1,2 \dots 2,0$ gute Ergebnisse liefern. Mit α_{fix} wird die generelle Geschwindigkeit der Adaption eingestellt. Die für die Schrittweitensteuerung im weiteren Verlauf dieser Arbeit verwendeten Parameter sind in Tab. 4.1 zusammengefasst. Die Glättungskonstante zur Bestimmung der geglätteten

Parameter	Wert
Minimale Schrittweite	$\alpha_{\min} = 0,0$
Maximale Schrittweite	$\alpha_{\max} = 2,0$
Fixe Schrittweite	$\alpha_{\text{fix}} = 0,3$

Tabelle 4.1: In dieser Arbeit verwendete Parametrierung der Schrittweitensteuerung.

Variablen $\overline{|X(\mu, k)|^2}$ und $\overline{|E(\mu, k)|^2}$ in Gl. (4.18) wurde auf $\beta = 0,3$ gesetzt, was einer schwachen Glättung entspricht. Eine größere Glättungskonstante würde zu einer starken Verschmierung der für die Nachhalldetektion wichtigen Signalfanken führen, wodurch diese nicht mehr präzise detektiert werden könnten.

4.3 Verstärkungsregelung

In Abschn. 4.1 wurde mit Gl. (4.14) die theoretisch optimale Schrittweite $\alpha_{\text{opt}}(\mu, k)$ hergeleitet. Der besseren Übersicht halber sei die Gleichung an dieser Stelle nochmals gezeigt

$$\alpha_{\text{opt}}(\mu, k) \approx \frac{|X(\mu, k)|^2}{|E(\mu, k)|^2} \cdot \|\mathbf{H}_{\Delta}(\mu, k)\|^2. \quad (4.23)$$

Zur Herleitung dieser Gleichung waren Näherungen notwendig, die nur gültig sind, wenn die Signale nicht korreliert sind. Diese Unkorreliertheit ist in einer geschlossenen elektroakustischen Schleife nicht gegeben.

Im vorangegangenen Abschn. 4.2 wurde eine nachhallbasierte Schrittweitensteuerung eingeführt. Diese basiert auf der Idee, dass im Nachhall eines rückgekoppelten Systems die Signale unkorreliert sind und das adaptive Filter somit ungestört gegen die korrekte Lösung konvergieren kann. Die Gl. (4.18) für die Nachhalldetektion sei ebenfalls hier nochmals gezeigt

$$\alpha_{\text{hall}}(\mu, k) = \frac{|X(\mu, k)|^2}{g^2(k) \cdot |E(\mu, k)|^2}. \quad (4.24)$$

Vergleicht man Gl. (4.23) mit Gl. (4.24) erkennt man, dass die nachhallbasierte Schrittweite $\alpha_{\text{hall}}(\mu, k)$ der optimalen Schrittweite $\alpha_{\text{opt}}(\mu, k)$ entspricht, wenn die Verstärkung $g^2(k)$ gleich dem inversen Systemabstand $\|\mathbf{H}_{\Delta}(\mu, k)\|^2$ ist, d. h.

$$\alpha_{\text{hall}}(\mu, k) = \alpha_{\text{opt}}(\mu, k) \quad (4.25)$$

wenn

$$g^2(k) = \frac{1}{\|\mathbf{H}_{\Delta}(\mu, k)\|^2}. \quad (4.26)$$

Das bedeutet, dass das Filter im Nachhall mit der optimalen Schrittweite konvergiert, wenn die Verstärkung vom inversen Systemabstand gesteuert wird. Da die Signale im Nachhall nicht korreliert sind, sind dann auch die Annahmen bzgl. der Unkorreliertheit der Signale zulässig, die bei der Herleitung der optimalen Schrittweite getroffen wurden. Während Sprachaktivität sind die Signale zwar nach wie vor korreliert, d. h. auch die Näherung ist nicht mehr gültig. Allerdings wird hier die Schrittweite durch die nachhallbasierte Schrittweitensteuerung verringert, sodass das Filter nicht bzw. nur langsam divergiert.

Dass die Verstärkung vom Systemabstand abhängt, hat einen weiteren Vorteil. Bei einer Raumänderung wird, wie in Abschn. 4.1.2 erläutert, der Systemabstand groß. In selbem Maße wird die Schrittweite erhöht, damit das Filter gegen die neue Impulsantwort konvergieren kann. Zusätzlich wird bei der Verstärkungsregelung die Verstärkung verringert, wodurch verhindert wird, dass das System instabil wird.

Bei der Betrachtung von Gl. (4.26) fällt auf, dass die Verstärkung breitbandig auf alle Frequenzen angewandt wird, während der Systemabstand frequenzabhängig ist. Diese Frequenzabhängigkeit liegt darin begründet, dass das adaptive Filter nur in Frequenzbändern abgleichen kann, in denen ein Anregungssignal vorhanden ist. Im Fahrzeug sind

dies vor allem die Frequenzbänder von Sprache und Fahrgeräusch, deren Spektren bereits in Abb. 2.15 auf S. 19 gezeigt worden sind. Zur Berechnung der Verstärkung wird daher der Systemabstand über diesen Frequenzbereich gemittelt

$$\tilde{g}(k) = \sqrt{\left(\frac{1}{\mu_o - \mu_u + 1} \sum_{\mu=\mu_u}^{\mu_o} \|\mathbf{H}_{\Delta}(\mu, k)\|^2 \right)^{-1}}, \quad (4.27)$$

wobei die Bandgrenzen zu $\mu_u = 2$ und $\mu_o = 93$ gewählt wurden, was bei einer Abtastfrequenz $f_s = 44,1$ kHz und einer DFT-Länge $N = 512$ einem Frequenzbereich von ca. 172 Hz bis 8010 Hz entspricht.

Die Verstärkung nach Gl. (4.27) würde mit kleiner werdendem Systemabstand immer weiter ansteigen. In einer realen Anwendung ist dies nicht sinnvoll, weshalb sie auf einen Maximalwert g_{\max} begrenzt wird, welcher vom Anwender eingestellt werden kann. Es folgt damit für die Verstärkungsregelung

$$g(k) = \begin{cases} \tilde{g}(k) & \text{wenn } \tilde{g}(k) < g_{\max}, \\ g_{\max} & \text{sonst.} \end{cases} \quad (4.28)$$

4.4 Schätzen des Systemabstands

Bei der im vorangegangenen Abschn. 4.3 vorgestellten nachhallbasierten Schrittweitensteuerung mit Verstärkungsregelung wird die Schrittweite nach Gl. (4.23) und die Verstärkung nach Gl. (4.28) berechnet. Für beide Berechnungen wird der Systemabstand $\|\mathbf{H}_{\Delta}(\mu, k)\|^2$ benötigt, welcher jedoch unbekannt ist. Aus diesem Grund sind die im Folgenden vorgestellten Verfahren zur Schätzung des Systemabstands notwendig. Bei diesen Schätzverfahren kann teilweise auf die Erkenntnisse aus der Echokompensation aufgebaut werden. So werden z. B. in [29, 46] zwei Verfahren erläutert, die im Kern sehr ähnlich zu den in dieser Arbeit untersuchten Verfahren sind. Bei dem Verfahren aus Abschn. 4.4.1 werden zur Bestimmung des Systemabstands Bereiche der Impulsantwort betrachtet, in denen angenommen wird, dass die reale Impulsantwort null ist. Das in Abschn. 4.4.2 erläuterte Verfahren basiert auf einem Minimumschätzer.

4.4.1 Nullen am Ende der Impulsantwort

In [29, Kap. 13.4.2.1] bzw. [46, Abschn. 4.2.1] werden Verfahren zur Systemabstandsschätzung mittels Totzeit-Koeffizienten beschrieben. Dabei wird das Lautsprechersignal *nach* der Abzweigung zum Kompensationsfilter $\hat{\mathbf{h}}(n)$ (vgl. Abb. 3.3 auf S. 27) künstlich verzögert. Dies hat zur Folge, dass der Kompensator die künstlich eingefügten Nullen ebenfalls schätzt. Da nun zumindest ein Teil der realen Impulsantwort, d. h. die eingefügten Nullen, bekannt ist, kann der Systemabstand für diesen Bereich berechnet werden.

Es wird dann davon ausgegangen, dass der Systemabstand über alle Koeffizienten gleichmäßig verteilt ist. Der im vorderen Bereich berechnete Systemabstand kann somit über die komplette Impulsantwort extrapoliert werden. Dieses Verfahren hat jedoch zwei Nachteile:

- Es wird eine künstliche Verzögerung eingefügt. Je genauer der Systemabstand berechnet werden soll, desto mehr Nullen müssen eingefügt werden. Bei der Echokompensation beispielsweise für Freisprecheinrichtungen kann diese zusätzliche Latenz tolerierbar sein. Im Echtzeitbetrieb eines ICC-Systems ist dies jedoch nicht umsetzbar.
- Das System friert ein, wenn sich die Impulsantwort des realen Systems ändert, da sich die künstlich eingefügten Nullen nicht mit ändern. Der geschätzte Systemabstand bleibt also fälschlicherweise klein.

In dieser Arbeit wurde die Idee des Verfahrens aufgegriffen und abgewandelt. Die zusätzliche Verzögerung wird umgangen, indem nicht die Nullen am Anfang der Impulsantwort zur Schätzung des Systemabstands verwendet werden, sondern der hintere Bereich, ab dem die Impulsantwort auf unter -60 dB abgefallen ist. Dazu werden die Messungen des vollbesetzten Fahrzeugs aus Abschn. 2.2, Abb. 2.10b herangezogen:

- Die Totzeit beträgt ca. $T_D = 5,6$ ms.
- Die Nachhallzeit ist ca. $T_{60} = 80,6$ ms.

Das bedeutet, dass nach etwa 86,2 ms die Energie unter -60 dB abgefallen ist. Bei einer Abtastfrequenz von $f_s = 44,1$ kHz entspricht dies etwa 3806 Abtastwerten, was bei einem Rahmenversatz von $L = 256$ wiederum etwa 14,8 Blöcke sind. Die blockweise Zerlegung

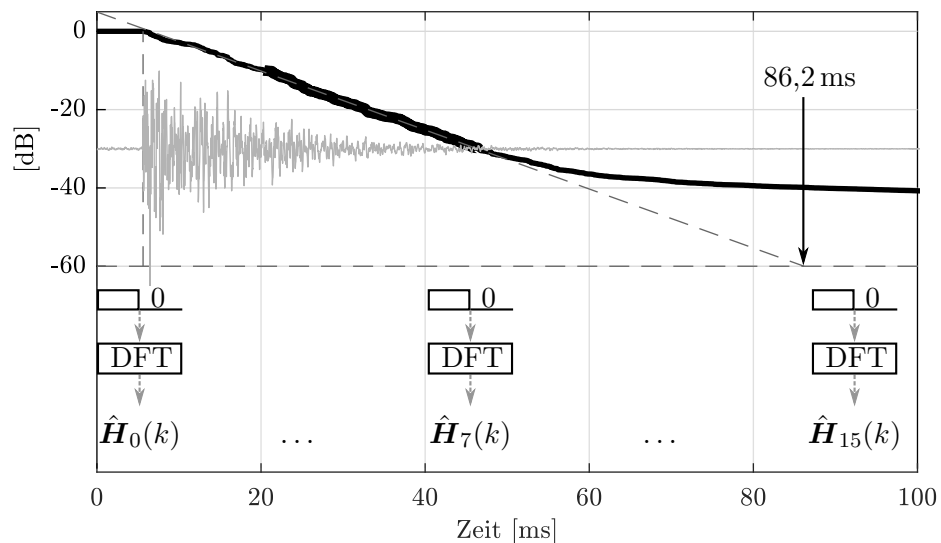


Abbildung 4.5: Zerlegung der zu schätzenden Impulsantwort in Teilfilter. Bei der gegebenen Blockgröße und Abtastfrequenz ist die Energie ab dem Teilfilter $\hat{H}_{15}(k)$ unter -60 dB abgefallen.

der Impulsantwort in ihre Teilfilter ist in Abb. 4.5 gezeigt. Aus der Abbildung wird

deutlich, dass ab dem Teilfilter $\hat{\mathbf{H}}_{15}(k)$ die Energie unter -60 dB abgefallen ist. Um die Impulsantwort bis einschließlich zum Teilfilter $\hat{\mathbf{H}}_{15}(k)$ zu schätzen, müssten in diesem Fall insgesamt $M = 16$ Teilfilter adaptiert werden. Wie in Abschn. 3.5.2 beschrieben, tragen die Teilfilter $\hat{\mathbf{H}}_8(k) \dots \hat{\mathbf{H}}_{14}(k)$ zur Kompensation nur wenig bei und werden für die Systemabstandsschätzung nicht benötigt. Um Rechenlast zu sparen, werden diese daher auf null gesetzt und nicht mitadaptiert

$$\hat{\mathbf{H}}_m(k) = \mathbf{0}_N, \quad m = 8, \dots, 14. \quad (4.29)$$

Hierbei ist $\mathbf{0}_N$ ein Nullvektor der Länge N . Es werden also insgesamt lediglich $16 - 7 = 9$ Teilfilter adaptiert.

Die Idee ist nun, den Systemabstand aus dem Abstand zwischen den Koeffizienten des geschätzten Teilfilters $\hat{\mathbf{H}}_{15}(k)$ und des realen Teilfilters $\mathbf{H}_{15}(k)$ zu extrapolieren, wobei man näherungsweise annimmt, dass an dieser Stelle die reale Impulsantwort bereits auf null abgeklungen ist, d. h.

$$\mathbf{H}_{15}(k) \approx \mathbf{0}_N \quad (4.30)$$

bzw.

$$H_{15}(\mu, k) \approx 0. \quad (4.31)$$

Damit erhält man für den frequenzselektiven Systemabstand näherungsweise

$$\overline{\|\mathbf{H}_\Delta(\mu, k)\|^2} \approx \gamma(\mu, k) = 9 \cdot \overline{|0 - \hat{H}_{15}(\mu, k)|^2}, \quad (4.32)$$

wobei $\gamma(\mu, k)$ den geschätzten Systemabstand bezeichnet. Wie beschrieben geht man davon aus, dass der Systemabstand über alle neun adaptierten Teilfilter gleichmäßig verteilt ist, weshalb mit dem Faktor neun multipliziert wird. Damit der geschätzte Systemabstand beim Wert eins startet, werden die Koeffizienten $\hat{H}_{15}(\mu, k = 0)$ entsprechend mit $\sqrt{1/9}$ initialisiert.

In der Simulation zeigt sich, dass es möglich ist, mit dem beschriebenen Verfahren den Systemabstand zu schätzen. Der oben genannte erste Nachteil, dass bei der Methode mit Totzeit-Koeffizienten eine zusätzliche Verzögerung eingefügt wird, kann durch die Berechnung des Systemabstands am Ende der Impulsantwort umgangen werden.

Es bleibt allerdings der zweite Nachteil, nämlich dass bei erfolgreichem Abgleich die Koeffizienten $\hat{H}_{15}(\mu, k)$ gegen null konvergieren und dort einfrieren. Der Grund ist, dass bei einer Raumänderung, wie sie z. B. durch Bewegungen der Passagiere verursacht wird, die Nachhallzeit sich nicht ändert. D. h. es ändern sich auch nur die Koeffizienten am Anfang der realen Impulsantwort, die Nullen am Ende bleiben bestehen. Das bedeutet, dass sich auch das adaptive Filter in diesem Bereich nicht ändert und damit der Systemabstand fälschlicherweise klein bleibt. Diese Problematik begründet die Notwendigkeit von Mechanismen zur Detektion von Raumänderungen.

4.4.2 Verfahren mit Minimumschätzer

Das in diesem Abschnitt vorgestellte Verfahren zur Schätzung des Systemabstands mittels Minimumschätzer ist ähnlich zu den Verfahren mit Kopplungsfaktor aus [29, Kap.

13.4.2.2]. Die dort beschriebenen Verfahren werden bei der schrittweitengesteuerten Echokompensation eingesetzt. Sie benötigen zusätzlich Mechanismen zur Detektion von Gegensprechen. Gegensprechen bedeutet beim Echokompensator, dass der ferne Teilnehmer und der lokale Sprecher gleichzeitig aktiv sind. In diesem Fall muss die Schrittweite verringert werden, da die Adaption durch den lokalen Sprecher gestört wird. Ein rückgekoppeltes System befindet sich während Sprachaktivität permanent in diesem Zustand des Gegensprechens. Einzelsprechen des fernen Teilnehmers bedeutet beim Echokompensator, dass es keine lokale Sprachaktivität gibt. In diesem Fall kann das adaptive Filter ungestört konvergieren. Bei der Rückkopplungskompensation entspricht dieser Zustand dem Nachhall. Mit dem im Folgenden beschriebenen Minimumschätzer kann die Einzelsprech- bzw. Nachhalldetektion direkt in den Schätzer integriert werden.

In Abschn. 4.1.1 wurde in den Gleichungen (4.9) und (4.11) gezeigt, dass der ungestörte Fehler $E_u(\mu, k)$ durch folgende Approximation näherungsweise berechnet werden kann

$$E\{|E_u(\mu, k)|^2\} \approx E\{|X(\mu, k)|^2\} \cdot E\{\|\mathbf{H}_\Delta(\mu, k)\|^2\}, \quad (4.33)$$

woraus direkt für den Systemabstand folgt

$$E\{\|\mathbf{H}_\Delta(\mu, k)\|^2\} \approx \frac{E\{|E_u(\mu, k)|^2\}}{E\{|X(\mu, k)|^2\}}. \quad (4.34)$$

Im realen System ist nur der gestörte Fehler $E(\mu, k)$ bekannt, $E_u(\mu, k)$ ist unbekannt. Allerdings gilt in Abwesenheit von lokaler Sprache $E_u(\mu, k) \approx E(\mu, k)$. Es wird also ein Verfahren zur Detektion von lokaler Sprachaktivität benötigt, um den Systemabstand nach Gl. (4.34) zu schätzen. Dies kann durch Verfolgen des Minimums des Verhältnisses $E(\mu, k)$ zu $X(\mu, k)$ realisiert werden. Der gestörte Fehler wird berechnet aus der additiven Überlagerung von ungestörtem Fehler und lokaler Sprache

$$E(\mu, k) = E_u(\mu, k) + S(\mu, k). \quad (4.35)$$

Daraus folgt, dass gelten muss

$$E\{|E_u(\mu, k)|^2\} \leq E\{|E(\mu, k)|^2\}. \quad (4.36)$$

Ersetzt man also in Gl. (4.34) den ungestörten Fehler $E_u(\mu, k)$ durch den gestörten Fehler $E(\mu, k)$ und folgt dem Minimum des zeitlichen Verlaufs erhält man näherungsweise den Systemabstand.

Das Verhältnis von Fehlersignal zu Anregungssignal wird mit $EXR(\mu, k)$ bezeichnet. Die Erwartungswerte werden, wie in Gl. (4.13) beschrieben, durch IIR-Glättung angenähert

$$EXR(\mu, k) = \frac{\overline{|E(\mu, k)|^2}}{\overline{|X(\mu, k)|^2}}. \quad (4.37)$$

D. h. der geschätzte Systemabstand $\gamma(\mu, k)$ ist

$$\gamma(\mu, k) = f_{\min}(EXR(\mu, k)) \approx \overline{\|\mathbf{H}_\Delta(\mu, k)\|^2}, \quad (4.38)$$

wobei $f_{\min}(\cdot)$ eine Funktion beschreibt, die dem Minimum des zeitlichen Verlaufs ihres Arguments folgt. Der Minimumschätzer wird rekursiv aktualisiert, indem die vorangegangene Schätzung mit $\varepsilon(\mu, k)$ gewichtet wird

$$\gamma(\mu, k) = \varepsilon(\mu, k) \cdot \gamma(\mu, k - 1). \quad (4.39)$$

Mit $\varepsilon(\mu, k)$ wird gesteuert, ob der Schätzer steigt oder fällt. Dabei ist $\varepsilon(\mu, k)$ wiederum abhängig vom Verhältnis von alter Schätzung zu aktuellem Eingang

$$\beta_{\min}(\mu, k) = \frac{\gamma(\mu, k - 1)}{EXR(\mu, k)}. \quad (4.40)$$

Folgende Zustände können dabei auftreten:

- $\beta_{\min}(\mu, k) > 1$: Die alte Schätzung ist größer als der aktuelle Eingang, d. h. der Schätzer muss fallen ($\varepsilon(\mu, k) = \varepsilon_f < 1$).
- $\beta_{\min}(\mu, k) \leq 1$: Die alte Schätzung ist kleiner oder gleich dem aktuellen Eingang. Lokale Sprachaktivität, der Schätzer muss auf dem alten Wert stehen bleiben, bzw. darf nur sehr langsam nach oben folgen ($\varepsilon(\mu, k) = \varepsilon_s \gtrsim 1$).

Zusätzlich wird ein Schwellwert β_{AP} eingeführt, mit welchem sich der Arbeitspunkt des Schätzers festlegen lässt. Für $\varepsilon(\mu, k)$ lässt sich dann schreiben

$$\varepsilon(\mu, k) = \begin{cases} \varepsilon_s & \text{wenn } \beta_{\min}(\mu, k) > \beta_{AP}, \\ \varepsilon_f & \text{wenn } \beta_{\min}(\mu, k) \leq \beta_{AP}. \end{cases} \quad (4.41)$$

Die durch Gl. (4.41) beschriebene Kennlinie ist in Abb. 4.6 dargestellt.

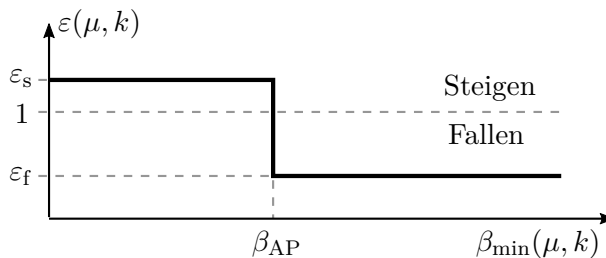


Abbildung 4.6: Kennlinie des Minimumschätzers.

Für Werte $\beta_{AP} > 1$ beginnt der Schätzer später zu fallen, d. h. erst, wenn die alte Schätzung um den Faktor β_{AP} größer ist als das aktuelle $EXR(\mu, k)$. Da in der Verstärkungsregelung der geschätzte Systemabstand invers in die Verstärkung eingeht, kann mit größeren Werten für β_{AP} der Anstieg der Verstärkung langsamer und dadurch das System gutmütiger eingestellt werden. Dies kann bei der Parametrierung des Algorithmus wichtig sein. In der Praxis hat sich ein Wertebereich von $1 \leq \beta_{AP} < 1,5$ als praktikabel herausgestellt. Die im weiteren Verlauf dieser Arbeit verwendeten Parameter sind in Tab. 4.2 gezeigt.

Parameter	Wert
Arbeitspunkt	$\beta_{\text{AP}} = 1,0$
Steigungskonstante	$\varepsilon_s = 1,001$
Abfallkonstante	$\varepsilon_f = 0,988$

Tabelle 4.2: In dieser Arbeit verwendete Parametrierung des Minimumschätzers.

Auch bei diesem Ansatz zeigt sich in der Simulation, dass der Minimumschätzer in der Lage ist, den tatsächlichen Systemabstand anzunähern. Aufgrund der kleinen Anzahl an einstellbaren Parametern sowie der geringen Komplexität ist das Verfahren zudem leicht zu handhaben.

Limitiert wird auch dieses Verfahren durch plötzlich auftretende Raumänderungen. In diesem Fall steigt hier zwar der Fehler $|E(\mu, k)|^2$ und damit auch $EXR(\mu, k)$ an. Allerdings kann der geschätzte Systemabstand diesem Anstieg aufgrund der geringen Steigungskonstante ε_s nicht schnell genug folgen und bleibt fälschlicherweise klein. Daher wird ein Verfahren benötigt, welches im Fall einer detektierten Raumänderung die Steigungskonstante auf einen größeren Wert setzt. Verschiedene Möglichkeiten dieses Problem zu lösen, werden im nächsten Abschnitt vorgestellt.

4.5 Detektion von Raumänderungen

Im vorangegangenen Abschnitt wurden zwei verschiedene Verfahren zur Schätzung des Systemabstands vorgestellt. Beide Verfahren sind in der Lage den Systemabstand korrekt zu schätzen, solange sich die Impulsantwort des zu schätzenden Raumes nicht ändert. Im Fall einer Raumänderung steigt der reale Systemabstand an. Damit die Schätzverfahren diesem Anstieg folgen können, werden Algorithmen zur Detektion von Raumänderungen benötigt. Im Folgenden werden verschiedene solcher Detektoren vorgestellt und ihre jeweiligen Vor- und Nachteile aufgezeigt.

Untersuchungen im Rahmen dieser Arbeit haben ergeben, dass sich die meisten der hier vorgestellten Detektoren sehr elegant mit dem in Abschn. 4.4.2 beschriebenen Minimumschätzer kombinieren lassen. Aus diesem Grund liegt das Hauptaugenmerk im weiteren Verlauf dieses Kapitels auf Raumänderungsdetektoren für den Minimumschätzer. Lediglich die in Abschn. 4.5.1 beschriebenen Schattenfilter können bei beiden Systemabstandsschätzern als Raumänderungsdetektoren zum Einsatz kommen. Allerdings zeigt sich, dass diese in einem geschlossenen System sehr schwer zu handhaben sind.

4.5.1 Schattenfilter

Schattenfilter sind bei der Echokompensation eine gängige Methode zur Detektion von Raumänderungen. Verschiedene Ansätze werden beispielsweise in [46, 35] beschrieben. Die Idee dieses Ansatzes ist es, ein Schattenfilter $\hat{h}_s(n)$ parallel zum Kompensationsfilter (Referenzfilter) adaptieren zu lassen. Das Prinzip ist in Abb. 4.7 gezeigt. Der Anschaulichkeit halber sind die Signale im Zeitbereich dargestellt. Das Schattenfilter hat eine

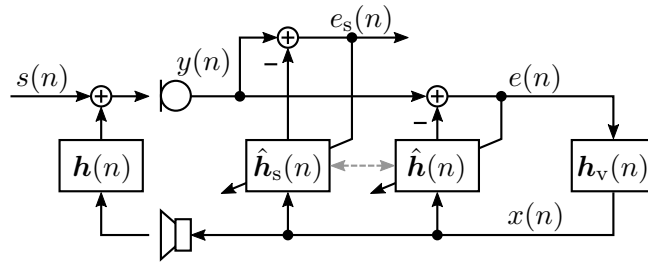


Abbildung 4.7: Schattenfilter zur Detektion von Raumänderungen.

Schrittweitensteuerung, die nicht einfrieren kann, bzw. lediglich eine fixe Schrittweite. D.h. das Schattenfilter wird ständig aktualisiert und friert auch nach erfolgtem Abgleich nicht ein, wodurch es Raumänderungen in jedem Abgleichzustand folgen kann. Die Schrittweite ist insgesamt größer als die des Referenzfilters. Zudem schätzt das Schattenfilter einen kleineren Teil der Impulsantwort, d.h. es werden weniger Koeffizienten adaptiert. Dadurch soll gewährleistet werden, dass das Schattenfilter schneller konvergiert als das Referenzfilter. Dass diese schnelle Konvergenz zulasten des erreichbaren Systemabstands geht, spielt für die Detektionsmechanismen keine Rolle.

Aus dem Vergleich von Schattenfilter und Referenzfilter, bzw. der mittleren Leistungen der Fehlersignale $\overline{|e(n)|^2}$ und $\overline{|e_s(n)|^2}$, sollen dann Erkenntnisse über den Abgleichzustand des Referenzfilters gewonnen werden. Vorausgesetzt das Schattenfilter konvergiert, so steigt beispielsweise die Leistung des Fehlersignals $\overline{|e_s(n)|^2}$ bei einer Raumänderung weniger stark an als $\overline{|e(n)|^2}$, da das Schattenfilter der Raumänderung schneller folgen kann als das Referenzfilter.

Versucht man das Verfahren für die Rückkopplungskompensation zu verwenden, stellt man fest, dass es aufgrund der geschlossenen Schleife sehr schwierig ist, das Schattenfilter zu parametrieren. Wie beim Referenzfilter ist auch hier die Korrelation zwischen Fehlersignal und Anregungssignal das Problem. Die Konvergenz des Schattenfilters ist ohne zusätzliche Dekorrelationsmaßnahmen nicht ausreichend, sodass Raumänderungen nicht präzise detektiert werden können. Daher wurde dieses Verfahren nicht weiterverfolgt.

4.5.2 Anstieg des Fehlersignals

Bei diesem Ansatz wird das Verhältnis $\beta_{\min}(\mu, k)$ nach Gl. (4.40) zur Detektion von Raumänderungen herangezogen. Ist der aktuelle Eingang $EXR(\mu, k)$ größer als die alte Schätzung $\gamma(\mu, k - 1)$, wird dies als Sprachaktivität interpretiert, der Minimumschätzer folgt nur sehr langsam nach oben. Ist der aktuelle Eingang jedoch viel größer als die alte Schätzung, wird dies als Raumänderung interpretiert, da davon ausgegangen wird, dass in diesem Fall der Fehler deutlich stärker ansteigt. Zur Detektion wird daher eine zweite Schwelle $\beta_R < \beta_{AP}$ eingeführt. Fällt $\beta_{\min}(\mu, k)$ unter diese Schwelle, wird $\varepsilon(\mu, k)$ auf einen Wert $\varepsilon_{ss} \gg 1$ gesetzt, damit der Schätzer schnell nach oben folgen kann. Damit

wird Gl. (4.41) erweitert zu

$$\varepsilon(\mu, k) = \begin{cases} \varepsilon_f & \text{wenn } \beta_{\min}(\mu, k) > \beta_{AP} \\ \varepsilon_s & \text{wenn } \beta_R < \beta_{\min}(\mu, k) \leq \beta_{AP} \\ \varepsilon_{ss} & \text{wenn } \beta_{\min}(\mu, k) \leq \beta_R. \end{cases} \quad (4.42)$$

Die dieser Fallunterscheidung zugehörige Kennlinie ist in Abb. 4.8 dargestellt, eine mögliche Parametrierung ist in Tab. 4.3 gezeigt.

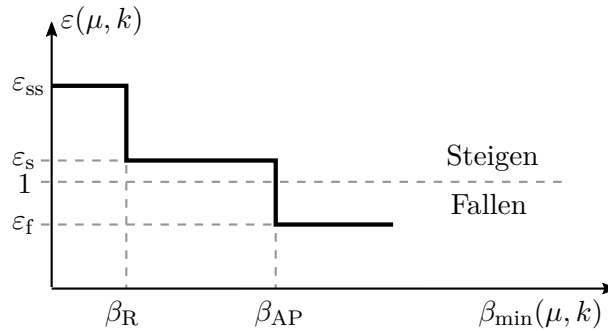


Abbildung 4.8: Kennlinie des Minimumschätzers mit zusätzlicher Schwelle zur Detektion von Raumänderungen. Fällt $\beta_{\min}(\mu, k)$ unter die Schwelle β_R , wird von einer Raumänderung ausgegangen.

Parameter	Wert
Schwelle schneller Anstieg	$\beta_R = 0,2$
Steigungskonstante schnell	$\varepsilon_{ss} = 1,02$

Tabelle 4.3: Beispielhafte Parametrierung des Verfahrens zur Detektion von Raumänderungen mit zusätzlicher Schwelle. Die weiteren Parameter wurden entsprechend Tab. 4.2 gewählt.

Ein Problem bei diesem Verfahren ist, dass aufgrund der geschlossenen Schleife das Anregungssignal $X(\mu, k)$ dem zeitlich verzögerten Fehlersignal $E(\mu, k)$ entspricht. D. h. bei einer Raumänderung wird der Fehler zwar groß, allerdings steigt kurz verzögert auch das Anregungssignal in gleichem Maße an. Das bedeutet, dass $EXR(\mu, k)$ nicht stark genug ansteigt und $\beta_{\min}(\mu, k)$ oftmals nicht unter die Schwelle β_R fällt. Setzt man die Schwelle β_R auf einen größeren Wert, werden zwar Raumänderungen präziser detektiert. Allerdings kommt es dann vor, dass auch während lokaler Sprache $\beta_{\min}(\mu, k)$ kleiner als β_R wird und dies als Raumänderung interpretiert wird. Als Folge steigt der geschätzte Systemabstand an, wodurch die Schrittweite fälschlicherweise groß wird. Zudem wird die Verstärkung nach unten geregelt bzw. konvergiert langsamer gegen den gewünschten Maximalwert.

In der Praxis kommt bei einer aktiven Geräuschreduktion ein weiteres Problem hinzu. Das Fehlersignal $E(\mu, k)$ durchläuft die Geräuschreduktion, bevor es als Lautsprecher-

gnal $X(\mu, k)$ wieder ausgegeben wird. Die Geräuschreduktion dämpft die geräuschbehafteten Bänder μ_b , weshalb die mittlere Leistung in diesen Bändern beim Lautsprecher-signal kleiner ist als beim Fehlersignal

$$|\overline{X(\mu_b, k)}|^2 < g^2(k) \cdot |\overline{E(\mu_b, k)}|^2. \quad (4.43)$$

Aufgrund der Geräuschreduktion ist $EXR(\mu, k)$ in den geräuschbehafteten Bändern folglich größer als in Teilbändern mit wenig Geräusch. Das Resultat ist, dass der Systemabstandsschätzer in diesen Bändern im Mittel sehr langsam fällt bzw. zeitweise sogar ansteigt, da $\beta_{\min}(\mu, k)$ häufig unter β_R sinkt.

4.5.3 Vergleich der hochpass- und tiefpassgefilterten Fehlerleistung

Bei diesem Verfahren wird das Leistungsdichtespektrum des Fehlersignals hochpass- und tiefpassgefiltert um Raumänderungen zu detektieren. Es wird davon ausgegangen, dass im Fall von lokaler Sprachaktivität der Fehler über den gesamten Frequenzbereich der Sprache gleichmäßig ansteigt. Im Gegensatz dazu steigt die Fehlerleistung bei einer Raumänderung hochfrequent stärker an als bei tiefen Frequenzen. Als Begründung hierfür wird in [47, 29] angegeben, dass die Wellenlänge tiefer Frequenzen groß ist im Vergleich zu den Abmessungen des Fahrzeuginnenraums bzw. zu den Bewegungen der Passagiere. Aus diesem Grund können Raumänderungen vor allem hochfrequent detektiert werden. Setzt man vor diesem Hintergrund die hochpass-/tiefpassgefilterten Fehlerleistungen ins Verhältnis, kann man die resultierende Größe zur Detektion von Raumänderungen verwenden.

Für die Bestimmung der tief-/hochfrequenten Fehlerleistung $\overline{P_{TP}(k)}$ und $\overline{P_{HP}(k)}$ werden die Teilbänder bis bzw. ab dem Teilband μ_g aufsummiert und mit einem IIR-Filter erster Ordnung geglättet. Die Glättungskonstante beträgt in diesem Fall $\beta = 0,9$

$$\overline{P_{TP}(k)} = \beta \cdot \overline{P_{TP}(k-1)} + (1 - \beta) \cdot \sum_{\mu=0}^{\mu_g-1} |E(\mu, k)|^2, \quad (4.44)$$

$$\overline{P_{HP}(k)} = \beta \cdot \overline{P_{HP}(k-1)} + (1 - \beta) \cdot \sum_{\mu=\mu_g}^{N-1} |E(\mu, k)|^2. \quad (4.45)$$

Übersteigt der Quotient

$$\beta_{HT}(k) = \frac{\overline{P_{HP}(k)}}{\overline{P_{TP}(k)}} \quad (4.46)$$

die Schwelle β_R , wird die Steuergröße $\varepsilon(\mu, k)$ des Minimumschätzers für alle Frequenzen $\mu = 0, \dots, N - 1$ auf einen Wert $\varepsilon_{ss}(k) \gg 1$ gesetzt. Damit wird die Schrittweite in sämtlichen Bändern gleichmäßig erhöht und die Verstärkung entsprechend herabgesetzt. Die Fallunterscheidung aus Gl. (4.41) ändert sich zu

$$\varepsilon(\mu, k) = \begin{cases} \varepsilon_f & \text{wenn } \beta_{\min}(\mu, k) > \beta_{AP} \text{ und } \beta_{HT}(k) \leq \beta_R \\ \varepsilon_s & \text{wenn } \beta_{\min}(\mu, k) \leq \beta_{AP} \text{ und } \beta_{HT}(k) \leq \beta_R \\ \varepsilon_{ss}(k) & \text{wenn } \beta_{HT}(k) > \beta_R \quad \forall \mu. \end{cases} \quad (4.47)$$

Es ist zu beachten, dass in der Fallunterscheidung in Gl. (4.47) in der unteren Zeile $\varepsilon(\mu, k)$ in *allen* Teilbändern μ auf den Wert $\varepsilon_{ss}(k)$ gesetzt wird. Hingegen werden in den oberen beiden Zeilen die Werte frequenzselektiv gesetzt.

Die Größe $\varepsilon_{ss}(k)$ läuft von einem großen Anfangswert $\varepsilon_{r0} \gg 1$ exponentiell auf den kleineren Wert $1 < \varepsilon_{r1} < \varepsilon_{r0}$ zurück

$$\varepsilon_{ss}(k) = \begin{cases} (\varepsilon_{ss}(k-1) - \varepsilon_{r1}) \cdot e^{-\frac{L}{f_s \cdot \tau_r}} + \varepsilon_{r1} & \text{wenn } \beta_{HT}(k) > \beta_R \\ \varepsilon_{r0} & \text{sonst.} \end{cases} \quad (4.48)$$

Hierbei ist τ_r die Zeitkonstante, innerhalb welcher $\varepsilon_{ss}(k)$ von ε_{r0} auf ε_{r1} fällt. Mit diesem Mechanismus kann der Systemabstandsschätzer sehr schnell auf Änderungen reagieren. Er wird jedoch unmittelbar mit der Zeitkonstante τ_r abgebremst, damit die Verstärkung nicht zu klein werden kann. Wäre dies nicht der Fall, würde es nach einem Eingriff sehr lange dauern, bis die gewünschte Maximalverstärkung wieder erreicht wird.

Ein Beispiel für die Parametrierung ist in Tab. 4.4 gezeigt. Bei der Bestimmung von μ_g

Parameter	Wert
Trennfrequenz	$\mu_g = 8$
Schwelle schneller Anstieg	$\beta_R = 30$
Anfangswert	$\varepsilon_{r0} = 1,4$
Endwert	$\varepsilon_{r1} = 1,001$
Zeitkonstante	$\tau_r = 0,11 \text{ s}$

Tabelle 4.4: Beispielhafte Parametrierung des Verfahrens zur Detektion von Raumänderungen durch Vergleich der hochpass-/tiefpassgefilterten Fehlerleistung. Die weiteren Parameter wurden entsprechend Tab. 4.2 gewählt.

wird aufgrund der eingeschränkten Bewegungsfreiheit der Passagiere davon ausgegangen, dass Bewegungen die Schallwellen bis zu einer Wellenlänge von 0,5 m beeinflussen. Bei einer Schallgeschwindigkeit von 340 m/s entspricht diese Wellenlänge einer Frequenz von ca. 680 Hz. Gerundet auf das nächste ganzzahlige Teilband bedeutet dies

$$\mu_g = \left\lfloor \frac{340 \text{ m/s}}{680 \text{ Hz}} \right\rfloor = 8. \quad (4.49)$$

Im realen Betrieb stellt sich jedoch heraus, dass es schwer ist, eine geeignete Trennfrequenz zu ermitteln. Ein Grund hierfür ist, dass z. B. die Stimmlage des Sprechers den Mechanismus beeinflusst. So kann es sein, dass ein Sprecher bzw. eine Sprecherin mit vielen hochfrequenten Stimmanteilen als Raumänderung detektiert wird. Heraufsetzen der Trennfrequenz führt dann allerdings dazu, dass tatsächliche Raumänderungen weniger präzise detektiert werden.

4.5.4 Vergleich der mittleren Leistungen von Mikrofonsignal und Fehlersignal

Bei diesem Verfahren wird eine Raumänderung anhand des Verhältnisses von Fehlersignal zu Mikrofonsignal detektiert. Aus dem Blockdiagramm des Rückkopplungskompen-

sators nach Abb. 3.9 auf S. 42 erkennt man, dass man das Fehlersignal $e(n)$ erhält, indem man die geschätzte Rückkopplung $\hat{r}(n)$ vom Mikrofonsignal $y(n)$ subtrahiert

$$e(n) = y(n) - \hat{r}(n) \quad \bullet \quad E(\mu, k) = Y(\mu, k) - \hat{R}(\mu, k). \quad (4.50)$$

Das Mikrofonsignal $y(n)$ wiederum setzt sich zusammen aus der lokalen Sprache $s(n)$ und dem Rückkopplungsanteil $r(n)$

$$y(n) = s(n) + r(n) \quad \bullet \quad Y(\mu, k) = S(\mu, k) + R(\mu, k). \quad (4.51)$$

Einsetzen von Gl. (4.51) in (4.50) führt auf

$$E(\mu, k) = S(\mu, k) + \underbrace{R(\mu, k) - \hat{R}(\mu, k)}_{=R_{\Delta}(\mu, k)} = S(\mu, k) + R_{\Delta}(\mu, k), \quad (4.52)$$

wobei $R_{\Delta}(\mu, k)$ die Differenz zwischen realer Rückkopplung $R(\mu, k)$ und geschätzter Rückkopplung $\hat{R}(\mu, k)$ bezeichnet. Bildet man nun das Betragsquadrat von Gl. (4.52), erhält man

$$|E(\mu, k)|^2 = |S(\mu, k) + R_{\Delta}(\mu, k)|^2 \quad (4.53)$$

$$= (S(\mu, k) + R_{\Delta}(\mu, k)) \cdot (S(\mu, k) + R_{\Delta}(\mu, k))^*. \quad (4.54)$$

Ausmultiplizieren führt auf

$$|E(\mu, k)|^2 = |S(\mu, k)|^2 + |R_{\Delta}(\mu, k)|^2 + S(\mu, k)R_{\Delta}^*(\mu, k) + S^*(\mu, k)R_{\Delta}(\mu, k). \quad (4.55)$$

Unter der Annahme, dass $S(\mu, k)$ und $R_{\Delta}(\mu, k)$ orthogonal sind, sind die Erwartungswerte der Kreuzterme null. Aufgrund der Korrelation zwischen Lautsprechersignal und lokaler Sprache beim rückgekoppelten System gilt diese Annahme allerdings nur in grober Näherung. Dennoch soll im Folgenden mit dieser Näherung weiter gerechnet werden. Man erhält damit für den Erwartungswert von Gl. (4.55)

$$\mathbb{E}\{|E(\mu, k)|^2\} = \mathbb{E}\{|S(\mu, k)|^2\} + \mathbb{E}\{|R_{\Delta}(\mu, k)|^2\}. \quad (4.56)$$

Über die gleiche Herleitung erhält man aus dem Mikrofonsignal nach Gl. (4.51)

$$\mathbb{E}\{|Y(\mu, k)|^2\} = \mathbb{E}\{|S(\mu, k)|^2\} + \mathbb{E}\{|R(\mu, k)|^2\}. \quad (4.57)$$

Der mittlere quadratische Fehler $\mathbb{E}\{|R_{\Delta}(\mu, k)|^2\}$ zwischen realer und geschätzter Rückkopplung ist immer dann groß, wenn das Filter schlecht abgeglichen ist. Ändert sich der Raum schneller als das Filter folgen kann, steigt $\mathbb{E}\{|R_{\Delta}(\mu, k)|^2\}$ proportional zum Systemabstand an. Da jedoch sowohl $R_{\Delta}(\mu, k)$, als auch $S(\mu, k)$ und $R(\mu, k)$ unbekannt sind, wird das Verhältnis der Leistungen von $Y(\mu, k)$ und $E(\mu, k)$ als Steuergröße zur Detektion von Raumänderungen verwendet. Setzt man dazu die Gleichungen (4.56) und (4.57) ins Verhältnis, erhält man

$$\frac{\mathbb{E}\{|E(\mu, k)|^2\}}{\mathbb{E}\{|Y(\mu, k)|^2\}} = \frac{\mathbb{E}\{|S(\mu, k)|^2\} + \mathbb{E}\{|R_{\Delta}(\mu, k)|^2\}}{\mathbb{E}\{|S(\mu, k)|^2\} + \mathbb{E}\{|R(\mu, k)|^2\}}. \quad (4.58)$$

Dieses Verhältnis wird größer eins, wenn gilt

$$\mathbb{E}\{|E(\mu, k)|^2\} > \mathbb{E}\{|Y(\mu, k)|^2\}. \quad (4.59)$$

In diesem Fall muss auch zwingend gelten

$$\mathbb{E}\{|R_\Delta(\mu, k)|^2\} > \mathbb{E}\{|R(\mu, k)|^2\}. \quad (4.60)$$

Steigt also die mittlere Fehlerleistung im laufenden Betrieb über die mittlere Mikrofonleistung, ist dies ein Indikator für eine Raumänderung.

Um das Verfahren implementieren zu können, werden auch hier die Erwartungswerte mittels IIR-Glättung erster Ordnung angenähert

$$\mathbb{E}\{|E(\mu, k)|^2\} \approx \overline{|E(\mu, k)|^2} = \beta \cdot \overline{|E(\mu, k-1)|^2} + (1-\beta) \cdot |E(\mu, k)|^2, \quad (4.61)$$

$$\mathbb{E}\{|Y(\mu, k)|^2\} \approx \overline{|Y(\mu, k)|^2} = \beta \cdot \overline{|Y(\mu, k-1)|^2} + (1-\beta) \cdot |Y(\mu, k)|^2, \quad (4.62)$$

wobei in diesem Fall die Glättungskonstante auf den Wert $\beta = 0,8$ gesetzt worden ist. Das Verhältnis

$$EYR(\mu, k) = \frac{\overline{|E(\mu, k)|^2}}{\overline{|Y(\mu, k)|^2}} \quad (4.63)$$

wird anschließend mit einem Spitzendetektor geglättet

$$\overline{EYR(\mu, k)} = \beta_{\text{peak}} \cdot \overline{EYR(\mu, k-1)} + (1-\beta_{\text{peak}}) \cdot EYR(\mu, k), \quad (4.64)$$

wobei β_{peak} gemäß folgender Fallunterscheidung

$$\beta_{\text{peak}} = \begin{cases} e^{-\frac{L}{\tau_r \cdot f_s}} & \text{wenn } \overline{EYR(\mu, k-1)} > EYR(\mu, k) \\ e^{-\frac{L}{\tau_a \cdot f_s}} & \text{sonst.} \end{cases} \quad (4.65)$$

berechnet wird. Die Zeitkonstanten τ_a und τ_r bezeichnen die Ansprech- und Rücklaufzeit des Spitzendetektors. Mit τ_a kann gesteuert werden, wie schnell $\overline{EYR(\mu, k)}$ einer Spitze nach oben folgen kann. Da sich gezeigt hat, dass eine möglichst schnelle Ansprechzeit gefordert ist, um bei einer Raumänderung schnell reagieren zu können, wird τ_a auf 0 s gesetzt. Mit τ_r wird gesteuert, wie schnell $\overline{EYR(\mu, k)}$ nach einer Spitze wieder abfällt.

Übersteigt $\overline{EYR(\mu, k)}$ die Schwelle β_R , wird schließlich die Steigungskonstante des Minimumschätzers auf einen Wert $\varepsilon(\mu, k) = \varepsilon_{\text{ss}} > 1$ gesetzt. Die Fallunterscheidung nach Gl. (4.41) wird damit zu

$$\varepsilon(\mu, k) = \begin{cases} \varepsilon_f & \text{wenn } \beta_{\text{min}}(\mu, k) > \beta_{\text{AP}} \text{ und } \overline{EYR(\mu, k)} \leq \beta_R, \\ \varepsilon_s & \text{wenn } \beta_{\text{min}}(\mu, k) \leq \beta_{\text{AP}} \text{ und } \overline{EYR(\mu, k)} \leq \beta_R, \\ \varepsilon_{\text{ss}} & \text{wenn } \overline{EYR(\mu, k)} > \beta_R. \end{cases} \quad (4.66)$$

Je größer die Rücklaufzeit τ_r , desto langsamer fällt $\overline{EYR(\mu, k)}$ zurück unter die Schwelle β_R , wodurch auch der Minimumschätzer länger ansteigen kann. Für untersuchten Szenarien hat sich die in Tab. 4.5 gezeigte Parametrierung als passend herausgestellt.

Parameter	Wert
Steigungskonstante schnell	$\varepsilon_{ss} = 1,02$
Schwelle schneller Anstieg	$\beta_R = 2,5$
Ansprechzeit	$\tau_a = 0,0 \text{ s}$
Rücklaufzeit	$\tau_r = 0,2 \text{ s}$

Tabelle 4.5: In dieser Arbeit verwendete Parametrierung des Verfahrens zur Detektion von Raumänderungen mittels Vergleich von Mikrofon- und Fehlerleistung. Die weiteren Parameter wurden entsprechend Tab. 4.2 gewählt.

Sowohl in der Simulation als auch im Demonstratorfahrzeug können mit diesem Verfahren Raumänderungen zuverlässig detektiert werden. Auch sehr drastische Raumänderungen, wie beispielsweise ein plötzlich vor den Lautsprecher gehaltenes Knie, werden ausreichend schnell detektiert, sodass das System auf diese Änderung reagieren kann. In den meisten Fällen kann sich der Kompensator an die neue Gegebenheit adaptieren, ohne dass sich das System aufschwingt.

4.6 Verfahren zur Detektion von Aufschwingen

Mit den im vorangegangenen Abschnitt beschriebenen Verfahren wird versucht, Raumänderungen frühzeitig zu detektieren, um in diesem Fall ein Aufschwingen zu verhindern. Dazu wird bei einer detektierten Raumänderung die Schrittweite des Rückkopplungskompensators erhöht, wodurch dieser der Impulsantwortänderung möglichst schnell folgen kann. Wird das System allerdings bei hohen Verstärkungen betrieben, kann es dennoch instabil werden, verursacht beispielsweise durch eine nicht- bzw. zu spät detektierte Raumänderung. Für diesen Fall wird ein Verfahren benötigt, welches das dadurch verursachte Aufschwingen detektiert und das adaptive Filter sowie die Verstärkung zurücksetzt, damit das System wieder in den stabilen Bereich geführt wird und sich neu einschwingen kann. Dieses Verfahren ist also eine Art Rettungsmechanismus, der immer dann eingreift, wenn der Rückkopplungskompensator nicht mehr konvergieren kann. Es ist lediglich als letzte Instanz zu verstehen, idealerweise wird das System in der Praxis so parametrierung, dass es gar nicht erst so weit kommt.

Der in Abschn. 4.5.4 vorgestellte Mechanismus wird aktiviert, sobald das mit einem Spitzendetektor geglättete Verhältnis von Fehlersignal zu Mikrofonsignal $\overline{EYR}(\mu, k)$ die Schwelle β_R übersteigt. In diesem Fall wird die Steigungskonstante des Minimumschätzers auf einen Wert $\varepsilon(\mu, k) = \varepsilon_{ss} > 1$ gesetzt.

Dieses Verfahren wird nun so erweitert, dass jedes Mal, wenn die Schwelle β_R überschritten wird, zusätzlich ein Zähler $\Delta_{ctr}(\mu, k)$ hochgezählt wird. Fällt $\overline{EYR}(\mu, k)$ unter die Schwelle β_R , wird der Zähler wieder auf null gesetzt

$$\Delta_{ctr}(\mu, k + 1) = \begin{cases} \Delta_{ctr}(\mu, k) + 1 & \text{wenn } \overline{EYR}(\mu, k) > \beta_R \\ 0 & \text{sonst.} \end{cases} \quad (4.67)$$

Anschließend wird kontrolliert, wie lange $\overline{EYR}(\mu, k)$ oberhalb der Schwelle bleibt. Dazu wird der Zähler gegen eine Zeitkonstante τ_{resc} verglichen

$$\Delta_{\text{resc}}(\mu, k) = \begin{cases} 1 & \text{wenn } \Delta_{\text{ctr}}(\mu, k) > \tau_{\text{resc}} \cdot \frac{f_s}{L} \\ 0 & \text{sonst.} \end{cases} \quad (4.68)$$

Die Variable $\Delta_{\text{resc}}(\mu, k)$ wird also auf den Wert eins gesetzt, wenn $\Delta_{\text{ctr}}(\mu, k)$ größer als die Zeitkonstante $\tau_{\text{resc}} \cdot f_s/L$ ist und auf null wenn sie kleiner ist. Wurde $\Delta_{\text{resc}}(\mu, k)$ in mehr als N_{resc} Teilbändern auf den Wert eins gesetzt, wird davon ausgegangen, dass das Filter divergiert ist. Deshalb werden schließlich sowohl das Filter $\hat{\mathbf{H}}(\mu, k)$ als auch der geschätzte der Systemabstand $\gamma(\mu, k)$ zurückgesetzt

$$\hat{\mathbf{H}}(\mu, k) = \begin{cases} \mathbf{0}_M & \text{wenn } \sum_{\mu=0}^{N-1} \Delta_{\text{resc}}(\mu, k) > N_{\text{resc}} \\ \hat{\mathbf{H}}(\mu, k) & \text{sonst.} \end{cases} \quad (4.69)$$

Hierbei bezeichnet $\mathbf{0}_M$ einen Nullvektor der Länge M . Durch zurücksetzen des geschätzten Systemabstands auf den Wert eins

$$\gamma(\mu, k) = \begin{cases} 1 & \text{wenn } \sum_{\mu=0}^{N-1} \Delta_{\text{resc}}(\mu, k) > N_{\text{resc}} \\ \gamma(\mu, k) & \text{sonst} \end{cases} \quad (4.70)$$

wird auch die vom Systemabstand gesteuerte Verstärkung auf 0 dB zurückgesetzt (vgl. Gl. (4.27)). Damit wird das System wieder stabilisiert und kann neu adaptieren. Die bei der Simulation verwendeten Parameter sind in Tab. 4.6 gezeigt.

Parameter	Wert
Anzahl Teilbänder	$N_{\text{resc}} = 128$
Zeitkonstante	$\tau_{\text{resc}} = 0,2 \text{ s}$

Tabelle 4.6: In dieser Arbeit verwendete Parametrierung des Rettungsmechanismus. Die weiteren Parameter wurden entsprechend Tab. 4.5 gewählt.

4.7 Ergebnisse und Simulationen

In diesem Abschnitt werden die Ergebnisse der in diesem Kapitel vorgestellten Algorithmen präsentiert. Dazu werden zwar in erster Linie die Simulationsergebnisse betrachtet, allerdings können die in den Simulationen gezeigten Effekte auch im realen Betrieb nachgewiesen werden. Bei den Simulationen werden zwei Szenarien getrennt untersucht:

In Abschn. 4.7.1 wird zunächst der Fall einer konstanten Impulsantwort betrachtet. An diesem Beispiel lassen sich vor allem die Funktionsweise der Schrittweitensteuerung, der Verstärkungsregelung sowie der Systemabstandsschätzung zeigen. Für letztere kommt dabei der in Abschn. 4.4.2 beschriebene Minimumschätzer zum Einsatz.

In Abschn. 4.7.2 wird anschließend eine Raumänderung simuliert und daran die Funktionsweise der Detektion einer Impulsantwortänderung und des Rettungsmechanismus gezeigt. Für die Raumänderungsdetektion werden dabei lediglich die Ergebnisse des in Abschn. 4.5.4 beschriebenen Verfahrens gezeigt, da sich dieses in der Praxis am besten bewährt hat.

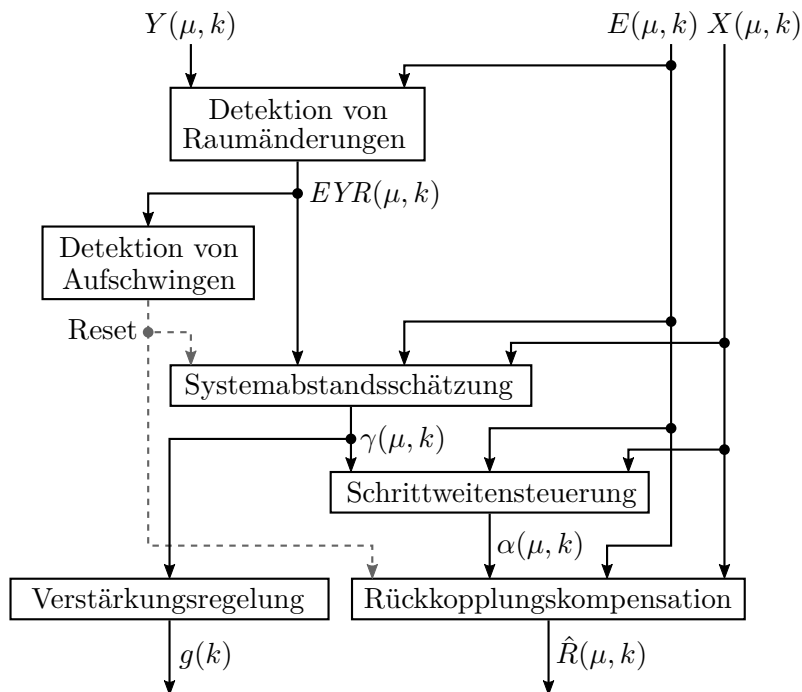


Abbildung 4.9: Blockdiagramm mit allen an der Rückkopplungskompensation beteiligten Teilsystemen.

Die an den nachfolgenden Simulationen beteiligten Teilsysteme und deren Abhängigkeiten untereinander sind in Abb. 4.9 als Blockdiagramm zusammenfassend dargestellt. Die benötigten Eingangsgrößen sind das Mikrofonsignal $Y(\mu, k)$, das Fehlersignal $E(\mu, k)$ und das Lautsprechersignal $X(\mu, k)$. Als Ausgangsgrößen erhält man die Verstärkung $g(k)$ sowie die geschätzte Rückkopplung $\hat{R}(\mu, k)$. Letztere wird zur Kompensation der Rückkopplung vom Mikrofonsignal abgezogen.

4.7.1 Simulation mit konstanter Impulsantwort

Die in diesem Abschnitt beschriebenen Simulationen werden mit der Impulsantwort des voll besetzten Fahrzeugs (vgl. Abb. 2.10b auf S. 14) durchgeführt. Als lokales Sprachsignal dient die Aufnahme eines männlichen Sprechers. Dieses Signal wurde im Stand mit dem Fahrermikrofon aufgezeichnet. Der Sprecher saß dabei auf dem Fahrersitz. Die Simulationsdauer beträgt ca. 19,5 s. Ein Hintergrundgeräusch wird nicht simuliert, was dem Fall einer idealen Geräuschreduktion entspricht. Das System wird also ausschließlich

mit Sprache angeregt. Ohne Rückkopplungskompensation wird die Stabilitätsgrenze der geschlossenen Schleife bei einer Verstärkung von 0 dB erreicht. Für die Simulation mit Rückkopplungskompensation wird die Zielverstärkung der Verstärkungsregelung nach Gl. (4.28) auf 50 dB eingestellt. Hierbei handelt es sich um einen beliebigen Wert, da mit der hier beschriebenen Schrittweitensteuerung das System auch für größere Verstärkungen stabil bleibt. Es muss an dieser Stelle zudem gesagt werden, dass eine derart hohe Verstärkung für den realen Betrieb eines ICC-Systems nicht realistisch ist. Es soll hier lediglich aufgezeigt werden, welcher enorme Gewinn mit den in dieser Arbeit entwickelten Algorithmen erzielt werden kann.

Zum Vergleich der maximalen stabilen Verstärkungen, die mit verschiedenen Ansätzen erreicht werden können, sind diese in Tab. 4.7 einander gegenüber gestellt. Verwendet man den Rückkopplungskompensator ohne Schrittweitensteuerung ist das System bis etwa 7 dB stabil. Der in [6] beschriebene Ansatz verwendet eine Schrittweitensteuerung ohne Verstärkungsregelung. Die Schrittweite wird hier direkt mit Gl. (4.18) gesteuert. Mit diesem Ansatz ist das System bis ca. 20 dB stabil. Die Verstärkungsregelung verbessert die Stabilität im Vergleich dazu nochmals bis hin zu Verstärkungen größer 50 dB.

Ohne Rückkopplungskompensation	0 dB
Rückkopplungskompensation mit fixer Schrittweite $\alpha(\mu, k) = 0,02$	≈ 7 dB
Nachhall-basierte Schrittweite ohne Verstärkungsregelung [6]	≈ 20 dB
Nachhall-basierte Schrittweite mit Verstärkungsregelung	> 50 dB

Tabelle 4.7: Die erreichbaren maximalen stabilen Verstärkungen mit verschiedenen Ansätzen.

Schrittweitensteuerung

Der zeitliche Verlauf der mit Gl. (4.22) gesteuerten Schrittweite ist in Abb. 4.10 gezeigt. Da die Schrittweite frequenzselektiv berechnet wird, ist sie in der Abbildung zur besseren Darstellung über den für Sprache relevanten Frequenzbereich von 125 Hz bis 8000 Hz gemittelt und anschließend geglättet. Man erkennt im oberen Teil der Abbildung, dass die Schrittweite ab ca. 9 s stark abfällt und von dort an klein bleibt. Dies liegt daran, dass die Schrittweite vom Systemabstand gesteuert wird. Je besser das System abgeglichen ist, desto kleiner wird also auch die Schrittweite. Hierdurch wird zwar die Konvergenzgeschwindigkeit langsamer, allerdings auch der Abgleich besser. Der untere Teil der Abbildung zeigt einen Detailausschnitt zwischen 14,0 s und 14,5 s. Dargestellt ist erneut die gemittelte Schrittweite, sowie zusätzlich das lokale Sprachsignal $s(n)$. Hier ist deutlich zu erkennen, dass die Schrittweite immer am Ende von Sprachaktivität, d. h. im Nachhall, stark ansteigt und beim Beginn bzw. während Sprachaktivität klein wird. So wird gewährleistet, dass der Kompensator nur dann konvergieren kann, wenn die Signale unkorreliert sind.

Die gestrichelte Linie bei 14,44 s markiert einen Zeitpunkt am Beginn von Sprachaktivität. Die Spektren von Schrittweite und lokaler Sprache zu diesem Zeitpunkt sind in

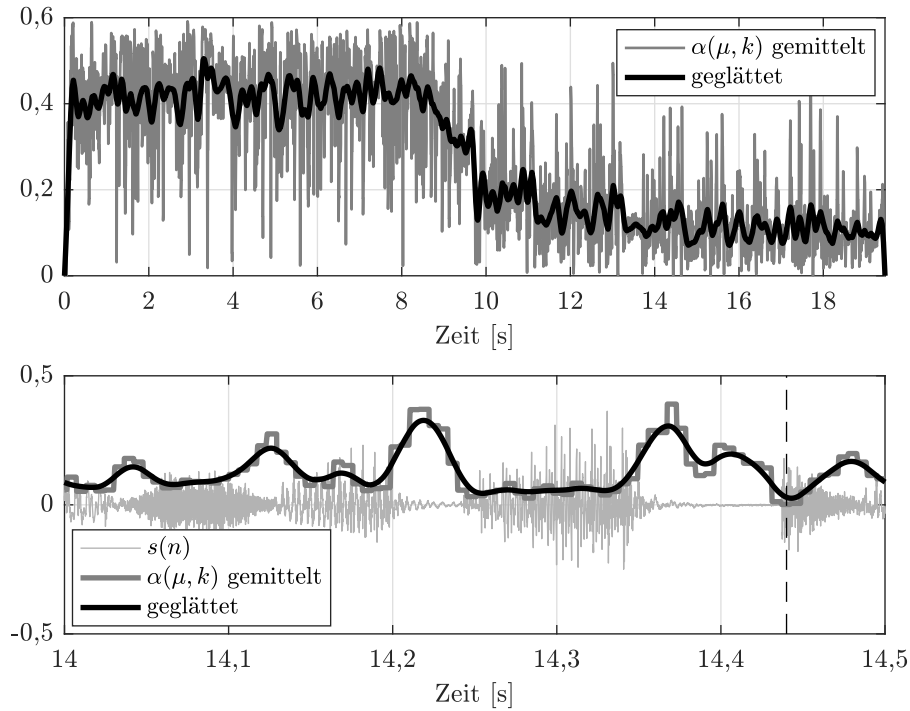


Abbildung 4.10: Der zeitliche Verlauf der Schrittweite. Im unteren Bild ist ein Detailausschnitt gezeigt. Die gestrichelte Linie bei 14,44s markiert Sprachaktivität (vgl. Abb. 4.11).

Abb. 4.11 zeigt. In dieser Abbildung wird das frequenzselektive Verhalten der Schritt-

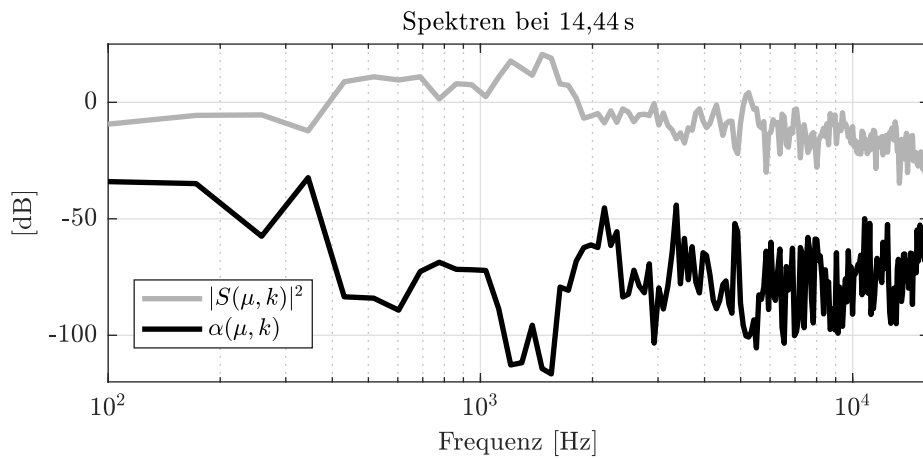


Abbildung 4.11: Die Spektren von Schrittweite und lokaler Sprache während Sprachaktivität. Man erkennt, dass die Schrittweite in Frequenzbändern mit viel Sprachanteilen klein ist.

weitensteuerung deutlich. In den Frequenzbereichen, in denen viele Sprachanteile vorhanden sind, ist zwangsweise auch die Korrelation groß. Damit der Kompensator in diesen Bändern nicht divergiert, muss die Schrittweite entsprechend klein sein. In der Abbildung ist dies z. B. im Bereich zwischen 1000 Hz und 2000 Hz sehr gut zu erkennen. Die beiden Überhöhungen im Sprachspektrum finden sich als deutliche Kerben in der Schrittweite wieder.

Systemabstandsschätzung und Verstärkungsregelung

Die für die Systemabstandsschätzung benötigten Größen sind in Abb. 4.12 dargestellt. Da diese Größen ebenfalls frequenzsektiv sind, wird exemplarisch der Verlauf eines einzelnen Teilbands $\mu = 4$ gezeigt. Die Frequenz dieses Teilbands liegt bei etwa 344,53 Hz und damit innerhalb des Sprachspektrums. In der Abbildung markiert die dunkelgraue Linie

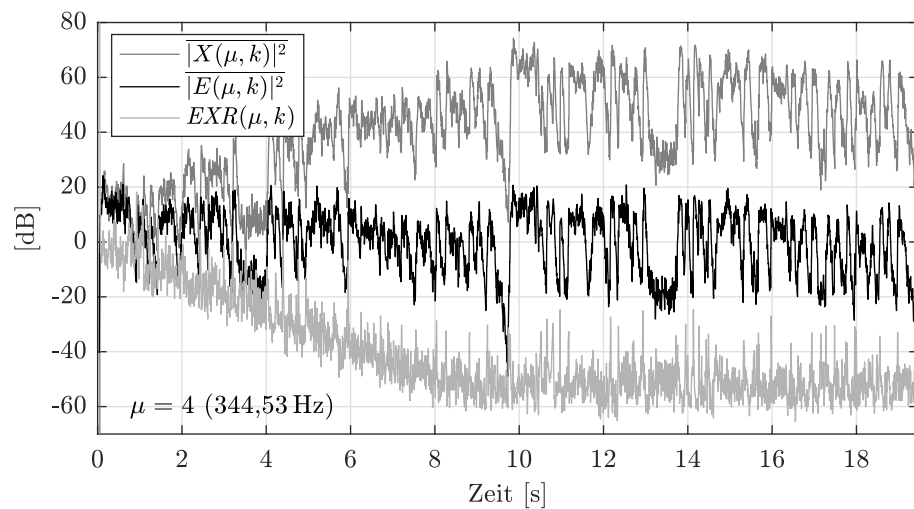


Abbildung 4.12: Der zeitliche Verlauf der mittleren Leistungen von Lautsprecher- und Fehlersignal, sowie des Verhältnisses $EXR(\mu, k)$ im Teilband $\mu = 4$.

(oben) die mittlere Leistung des Lautsprechersignals. Dass diese bis etwa 10 s ansteigt, liegt an der Verstärkungsregelung, die zu diesem Zeitpunkt ihr Maximum erreicht. Die Fehlerleistung (schwarz) wird hingegen nicht verstärkt und schwankt aus diesem Grund um 0 dB. Zur Bestimmung des Systemabstands werden die beiden Leistungen gemäß Gl. (4.37) ins Verhältnis gesetzt. Anschließend wird dem Minimum der resultierenden Größe $EXR(\mu, k)$ gefolgt.

Der Verlauf der Steuergröße $\beta_{\min}(\mu, k)$ des Minimumschätzers nach Gl. (4.40) ist in Abb. 4.13 gezeigt. Auch hier wird lediglich der Verlauf eines einzelnen Teilbands betrachtet. Zudem wurde in der Abbildung $\beta_{\min}(\mu, k)$ geglättet, um eine anschaulichere Darstellung zu erhalten. Wie aus der Fallunterscheidung nach Gl. (4.41) hervorgeht, fällt der Minimumschätzer, wenn $\beta_{\min}(\mu, k)$ oberhalb der Schwelle β_{AP} ist. In der Abbildung ist dies bis ca. 10 s der Fall. Danach liegt $\beta_{\min}(\mu, k)$ die meiste Zeit unterhalb der Schwelle.

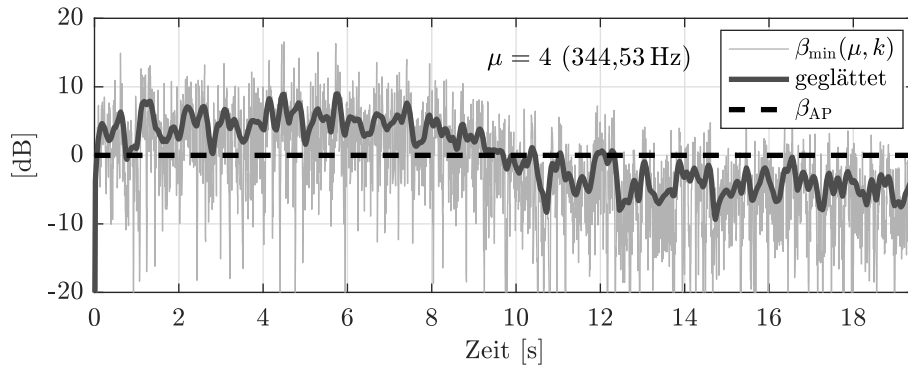


Abbildung 4.13: Der zeitliche Verlauf der Steuergröße des Minimumschätzers $\beta_{\min}(\mu, k)$ im Teilband $\mu = 4$.

Dies hat zur Folge, dass der geschätzte Systemabstand ab diesem Zeitpunkt nicht weiter bzw. nur noch sehr langsam sinkt.

Das Resultat der Systemabstandsschätzung ist schließlich im oberen Teil von Abb. 4.14 zu sehen. Erneut sind die Verläufe im Teilband $\mu = 4$ gezeigt. Es ist zu erkennen, dass der reale Systemabstand $\|\mathbf{H}_{\Delta}(\mu, k)\|^2$ näherungsweise dem Minimum des Verhältnisses $EXR(\mu, k)$ entspricht. Wie bereits nach der Betrachtung von Abb. 4.13 zu erwarten war, fällt der geschätzte Systemabstand $\gamma(\mu, k)$ (schwarz) bis ca. 10s und sinkt danach nur noch langsam. Man erkennt außerdem, dass der geschätzte Systemabstand in guter Näherung dem realen Systemabstand entspricht, allerdings liegt die Schätzung die meiste Zeit oberhalb des realen Systemabstands. Eine genauere Schätzung könnte in diesem Fall mit einer kleineren Abfallkonstante ε_f erreicht werden. In der Praxis hat sich jedoch gezeigt, dass es vorteilhaft sein kann, den Verlauf des realen Systemabstands zu überschätzen. Der Grund hierfür ist die Verstärkungsregelung. Diese ist im unteren Teil der Abb. 4.14 zu sehen. Dort ist die breitbandige Verstärkung $g(k)$ dargestellt, welche nach Gl. (4.27) aus dem über den Sprachbereich gemittelten geschätzten Systemabstand berechnet wird. Dementsprechend entspricht der inverse Verlauf der Verstärkung bis ca. 8s näherungsweise dem im oberen Bild gezeigten geschätzten Systemabstand. Würde der geschätzte Systemabstand schneller abfallen, würde demnach auch die Verstärkung schneller ansteigen. Dies könnte wiederum zur Folge haben, dass es dem Kompensator nicht gelingt, ausreichend schnell zu konvergieren, was zwangsläufig zu Instabilität führen würde. Schließlich zeigt die Abbildung, dass nach Erreichen der maximalen Verstärkung bei ca. 8,5s die Verstärkung auf diesem Wert festgehalten wird. Der Systemabstand sinkt ab diesem Zeitpunkt nur noch langsam.

4.7.2 Simulation von Raumänderungen

Die im vorangegangenen Abschnitt beschriebenen Verstärkungen von über 50 dB lassen sich nur dann realisieren, wenn die Impulsantwort zwischen den Lautsprechern und dem Mikrofon konstant ist. Ändert sich bei einer sehr großen Verstärkung der Raum, schwingt sich das System sofort auf, da der Kompensator nicht schnell genug auf die Änderung

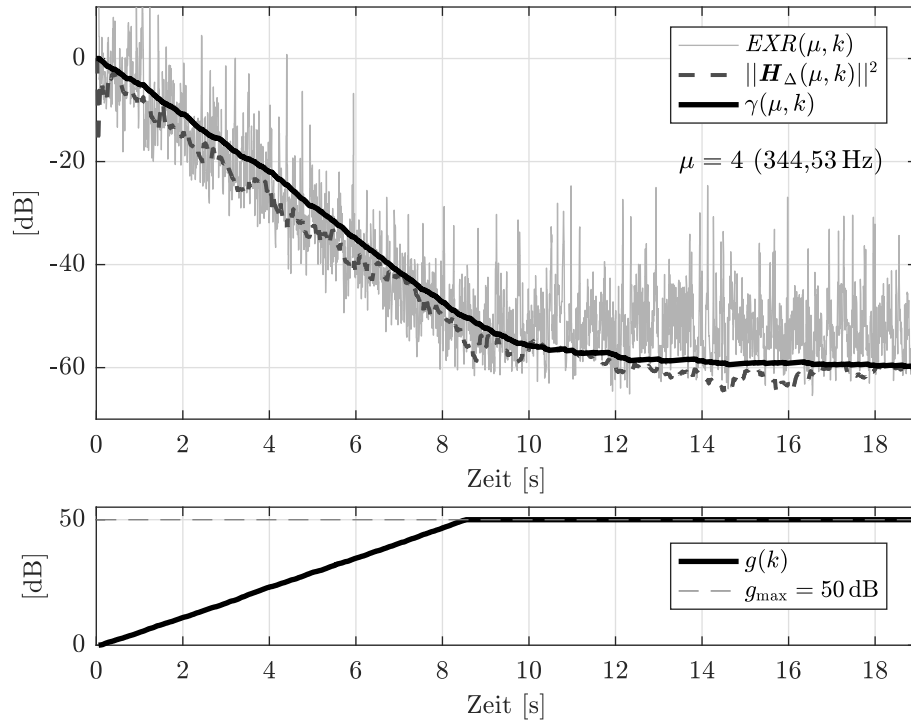


Abbildung 4.14: Der zeitliche Verlauf von realem und geschätztem Systemabstand im Teilband $\mu = 4$, sowie der Verlauf der Verstärkung. Man erkennt, dass die Verstärkung aufgrund der Verstärkungsregelung invers zum Systemabstand verläuft.

reagieren kann. Im realen Betrieb finden diese Impulsantwortänderungen aufgrund von Bewegungen der Passagiere ständig statt. Kleine Änderungen werden dabei mit dem in Abschn. 4.5.4 beschriebenen Verfahren detektiert. Ist dieses Verfahren nicht mehr in der Lage, ein Aufschwingen zu verhindern, greift der in Abschn. 4.6 beschriebene Rettungsmechanismus. Die Impulsantwortänderungen werden im Folgenden durch Überblenden zweier verschiedener Impulsantworten simuliert.

Detektion von Raumänderungen

Als lokales Sprachsignal dient erneut die oben beschriebene Aufnahme im Stand. Im ersten Schritt wird die Zielverstärkung auf 10 dB eingestellt. Nach erfolgter Konvergenz des Rückkopplungskompensators ändert sich die zu schätzende Impulsantwort. Hierbei wird das Szenario simuliert, dass der Passagier hinten links sein Knie vor den seitlichen Lautsprecher hält. Nach 10 s nimmt er das Knie plötzlich weg (vgl. auch Abb. 2.11 auf S. 15). In der Simulation werden dafür die beiden Impulsantworten innerhalb 0,5 s linear ineinander übergeblendet.

Die resultierenden zeitlichen Verläufe der Größen sind in Abb. 4.15 gezeigt. Da sich in diesem Fall die Raumänderung vor allem in den höheren Frequenzen auswirkt, ist hier

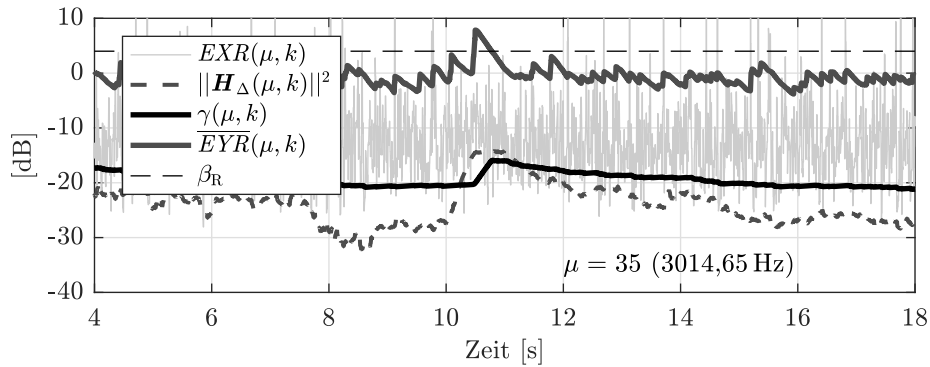


Abbildung 4.15: Der zeitliche Verlauf der für die Raumänderungsdetektion benötigten Größen. Die Raumänderung findet hier bei 10 s statt. Dargestellt sind die Signale im Teilband $\mu = 35$.

exemplarisch der Verlauf des Teilbands $\mu = 35$ bzw. 3014,65 Hz gezeigt. Zu sehen ist, dass der reale Systemabstand $\|\mathbf{H}_\Delta(\mu, k)\|^2$ aufgrund der Raumänderung bei 10 s stark ansteigt. Diesem Anstieg muss der geschätzte Systemabstand folgen. Nach Gl. (4.41) wird zu diesem Zweck die Steigungskonstante des Minimumschätzers erhöht, wenn das Verhältnis $\overline{EYR}(\mu, k)$ über die Schwelle β_R steigt. In der Abbildung ist zu sehen, dass dies unmittelbar nach der Impulsantwortänderung der Fall ist. Entsprechend steigt auch der geschätzte Systemabstand $\gamma(\mu, k)$ schneller an. Durch die Erhöhung der Schrittweite kann der Rückkopplungskompensator gegen die neue Impulsantwort konvergieren. Dies äußert sich darin, dass ab ca. 11 s sowohl der reale als auch der geschätzte Systemabstand erneut zu sinken beginnen.

Detektion von Aufschwingen

Bei diesem zweiten Szenario kommt wieder das gleiche Sprachsignal wie oben zum Einsatz. Auch die Zielverstärkung wird abermals auf 10 dB eingestellt. Um das System zum Aufschwingen zu bringen, wurde in diesem Fall jedoch eine stärkere Impulsantwortänderung simuliert. Bis 10 s hält der Fahrer seine Hand vor das Mikrofon, dann nimmt er sie plötzlich weg. In der Simulation werden dafür die beiden Impulsantworten aus Abb. 2.12 auf S. 15 innerhalb 0,5 s ineinander überblendet. Wie aus der Abbildung ersichtlich ist, dämpft die vorgehaltene Hand nahezu das komplette Spektrum zw. 5 dB und 10 dB. Nimmt man diese Dämpfung durch Entfernen der Hand plötzlich weg, schwingt sich das System auf, weshalb der in Abschn. 4.6 beschriebene Rettungsmechanismus benötigt wird.

In Abb. 4.16 ist zu sehen, dass in der Folge der Impulsantwortänderung der geschätzte Systemabstand ab ca. 10,3 s zu steigen beginnt. Hier wurde durch das in Abschn. 4.5.4 beschriebene Verfahren die Steigungskonstante auf ε_{ss} gesetzt. Da diese Maßnahme das System nicht ausreichend schnell stabilisiert, greift ab 10,5 s der Rettungsmechanismus und setzt das Filter sowie den geschätzten Systemabstand zurück. Durch diesen Mechanismus wird auch die im unteren Teil der Abbildung zu sehende Verstärkung zurückgesetzt.

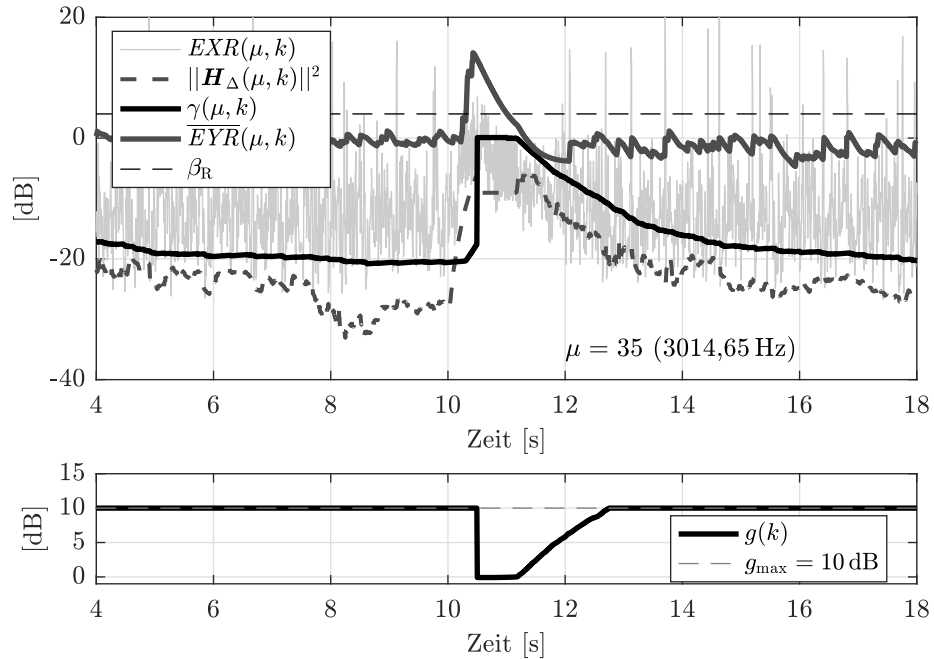


Abbildung 4.16: Der zeitliche Verlauf von geschätztem Systemabstand, der für die Schätzung benötigten Größen im Teilband $\mu = 35$ sowie der Verstärkung. Erneut liegt bei 10s eine Raumänderung vor.

Sobald $\overline{EYR}(\mu, k)$ bei ca. 11 s wieder unter die Schwelle β_R sinkt, konvergieren die Filterkoeffizienten gegen die neue Impulsantwort und die Verstärkung wird wieder auf 10 dB nach oben geregelt.

Der Verlauf der quadrierten Betragsfrequenzgänge ist in Abb. 4.17 zu sehen. Für die Darstellung werden die Betragsquadrate der M Teilfilter gemittelt. Im oberen Teil der Abbildung ist bei 10s die simulierte Raumänderung zu sehen. Im unteren Teil der Abbildung erkennt man, dass bei 10,5s die Filterkoeffizienten zurückgesetzt werden und das Filter ab etwa 11,1s wieder neu konvergiert.

4.8 Zusammenfassung

In diesem Kapitel wurde eine neuartige Schrittweitensteuerung für die adaptive Rückkopplungskompensation vorgestellt. Bei dieser Schrittweitensteuerung wird der Nachhall ausgenutzt, da hier die Signale nicht korreliert sind und das adaptive Filter somit ungestört konvergieren kann. Es wurde gezeigt, dass die Nachhall-basierte Schrittweitensteuerung als optimale Schrittweitensteuerung interpretiert werden kann, wenn die Verstärkung im Vorwärtspfad des ICC-Systems mit dem inversen Systemabstand geregelt wird. Sowohl für die Schrittweitensteuerung als auch für die Verstärkungsregelung muss der Systemabstand bekannt sein. Aus diesem Grund wurden zwei verschiedene Verfahren zum Schätzen des Systemabstands vorgestellt. Die große Herausforderung hierbei

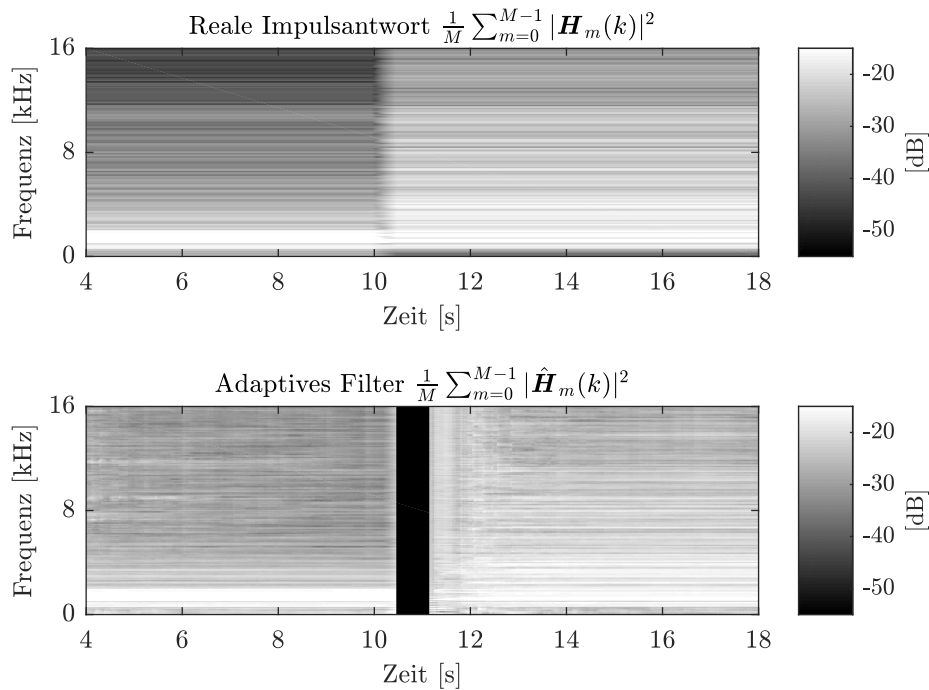


Abbildung 4.17: Der zeitliche Verlauf der quadrierten Betragsfrequenzgänge von realer Impulsantwort und adaptivem Filter. Oben ist die Impulsantwortänderung bei 10 s deutlich zu erkennen.

ist es, Raumänderungen zu detektieren. Der Grund dafür ist, dass der reale Systemabstand im Fall einer Raumänderung schnell ansteigen kann und der Schätzer dieser Änderung folgen muss. Zu diesem Zweck wurden vier verschiedene Algorithmen zur Detektion von Raumänderungen beschrieben. Wenn es nicht gelingt, eine Raumänderung präzise zu detektieren, kann das System instabil werden. Um dies zu verhindern, wurde ein Mechanismus entwickelt, welcher ein Aufschwingen detektiert und die Algorithmen sowie die Verstärkung des Systems zurücksetzt. In den Simulationen am Ende des Kapitels zeigte sich, dass das System mit dem beschriebenen Verfahren auch bei hohen Verstärkungen stabil bleibt. Weil das Verfahren im Frequenzbereich implementiert wird und Eigenschaften der Filterbank ausnutzt, ist es zudem recheneffizient.

Kapitel 5

Postfilter zur Reduktion der Restrückkopplung

Im praktischen Betrieb spielt die für die Algorithmen benötigte Rechenleistung eine große Rolle. Einen erheblichen Anteil an der Rechenlast hat die Rückkopplungskompensation. Dies kann insbesondere bei großen Filterlängen kritisch werden. Wie bereits erläutert, kann das adaptive Filter aus diesem Grund im praktischen Einsatz nur einen begrenzten Teil der realen Impulsantwort abdecken, wodurch die späten Anteile der Impulsantwort nicht kompensiert werden. Diese späten Anteile können einen hörbaren Nachhall verursachen, welcher den Raum halliger erscheinen lässt und das klangliche Ergebnis dadurch beeinträchtigt. Beinhalten die späten Anteile Reflexionen mit großer Amplitude, kann sich das System im schlechtesten Fall sogar trotz abgeglichenem Kompensator aufschwingen. Äquivalent zum Restecho bei der Echokompensation werden diese nicht kompensierten Rückkopplungen im Folgenden als Restrückkopplung bezeichnet.

Zur Reduktion des Restechos werden bei der Echokompensation Postfilter eingesetzt. Diese Postfilter werden oftmals mit Wiener-Filtern oder mit spektraler Subtraktion realisiert. Ein Ansatz ist es, das LDS des Restechos, ähnlich wie auch die optimale Schrittweite, aus dem Verhältnis von ungestörter zu gestörter Fehlerleistung zu schätzen [28, 15].

Ein ähnliches Verfahren wird in diesem Kapitel zur Reduktion der Restrückkopplungen vorgestellt. Im Unterschied zu den oben genannten Verfahren zur Restechoreduktion wird hier jedoch die Restrückkopplung mit einem rekursiven Hallmodell geschätzt. Anschließend wird sie mit einem Wiener-Filter, wie in Abschn. 3.4 beschrieben, spektral unterdrückt. Die Basis zur Schätzung des LDS der Restrückkopplung stellt dabei die im nachfolgenden Abschnitt vorgestellte modellbasierte Rückkopplungsreduktion dar. Diese wird in den darauf folgenden Abschnitten so modifiziert, dass sie als Postfilter für die Rückkopplungskompensation eingesetzt werden kann. Die Ergebnisse finden sich im letzten Abschnitt. Teile dieses Kapitels wurden bereits in [24] veröffentlicht.

5.1 Modellbasierte Rückkopplungsreduktion

Die Idee der in diesem Abschnitt vorgestellten modellbasierten Rückkopplungsreduktion basiert auf Verfahren zur Enthaltung von Sprachsignalen mit Hilfe eines statistischen Modells der Impulsantwort, wie z. B. in [40, 27] beschrieben. Bei diesen Verfahren wird die für den Nachhall verantwortliche Impulsantwort als nicht-stationärer, gaußverteilter Zufallsprozess modelliert. Mit dem Modell der Impulsantwort kann das LDS des Nachhalls geschätzt werden. Die Ansätze zur Enthaltung werden von dem bereits in Abschn. 3.2 erwähnten Verfahren [78] aufgegriffen und zur Reduktion der Rückkopplungsanteile am Mikrofon eines ICC-Systems angewandt. Das LDS der Rückkopplung

wird in [78] mit einem rekursiven Modell geschätzt, welches aus der vorab gemessenen Impulsantwort abgeleitet wird. In dem zitierten Artikel wird das Verfahren als alleinige Maßnahme zur Stabilisierung eines ICC-Systems vorgestellt, eine Kombination mit einer Rückkopplungskompensation ist nicht vorgesehen. Auch in diesem Abschnitt soll die modellbasierte Rückkopplungsreduktion zunächst ohne Rückkopplungskompensation vorgestellt werden.

In Abb. 5.1 ist das entsprechende Blockdiagramm gezeigt. Ohne Kompensator entspricht das Fehlersignal des Kompensators $e(n)$ dem Mikrofonsignal $y(n)$ (vgl. auch Abb. 3.7 auf S. 36).

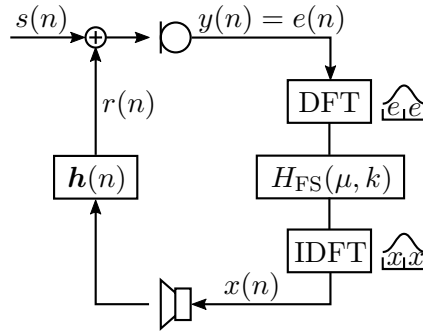


Abbildung 5.1: Blockdiagramm der modellbasierten Rückkopplungsreduktion mit Wiener-Filter. Die Rückkopplungskompensation wird zunächst nicht mit betrachtet.

Die Übertragungsfunktion $H_{FS}(\mu, k)$ des Filters zur Rückkopplungsreduktion ist eine Approximation des optimalen Wiener-Filters, welches in Abschn. 3.4.1 in Gl. (3.56) hergeleitet wurde zu

$$H_{w,opt}(\mu, k) = 1 - \frac{P_{RR}(\mu, k)}{P_{EE}(\mu, k)}. \quad (5.1)$$

Wie bereits erläutert, liegt hier die Annahme zugrunde, dass $r(n)$ und $s(n)$ orthogonal sind.

Da die Leistungsdichtespektren $P_{RR}(\mu, k)$ und $P_{EE}(\mu, k)$ nicht bekannt sind, müssen sie geschätzt werden. Im Fall der Rückkopplung $P_{RR}(\mu, k)$ wird hierfür ein rekursives Modell des Übertragungspfades zwischen Lautsprecher und Mikrofon entwickelt. Mit diesem ist es möglich, anhand des Lautsprechersignals abzuschätzen, welche Anteile davon über den Raum zum Mikrofon zurück gekoppelt werden. Der Übertragungspfad wird durch die in Abschn. 2.2 erläuterten Größen Totzeit T_D , Nachhallzeit T_{60} sowie dem Betragsfrequenzgang beschrieben. Ein am Lautsprecher abgestrahltes Signal kommt um die Totzeit T_D verzögert am Mikrofon an. Aufgrund der frequenzabhängigen Dämpfung des Raumes erfährt das Signal eine spektrale Färbung. Diese wird durch die Kopplungsfaktoren $A(\mu)$ beschrieben. Nach Abschalten der Schallquelle klingt das Signal exponentiell mit der Nachhallzeit T_{60} ab. Diese drei Parameter werden einmalig vorab aus der im Fahrzeug gemessenen Impulsantwort bestimmt. Zur Vereinfachung werden in der vorlie-

genden Arbeit zwei Näherungen getroffen:

1. Bei der Totzeit kann es vorkommen, dass einzelne Frequenzbänder über den Körperschall schneller übertragen werden als durch den Luftschall. Zudem dämpfen die im Fahrzeug verbauten Materialien verschiedene Frequenzen unterschiedlich stark, weshalb auch die Nachhallzeit über die Frequenz schwankt. Dennoch hat sich herausgestellt, dass das Modell hinreichend genau ist, wenn Totzeit und Nachhallzeit über den kompletten Frequenzbereich als konstant angenommen werden.
2. In Abschn. 2.2 wurde erläutert, dass sich die Raumeigenschaften ändern können. Dies wird hier nicht berücksichtigt, da sich gezeigt hat, dass ein statisches Modell für das Verfahren ausreichend ist. Aus diesem Grund wird das Modell einmalig aus einer vorab im leeren Fahrzeug gemessenen Impulsantwort bestimmt.

Totzeit und Nachhallzeit werden wie in Abschn. 2.2 beschrieben aus der Energieabklingkurve bestimmt. Da der Algorithmus blockweise im Frequenzbereich berechnet wird, müssen die Parameter umgerechnet werden. Die Totzeit umgerechnet in Blöcke ist

$$K = \left\lceil T_D \cdot \frac{f_s}{L} \right\rceil, \quad (5.2)$$

wobei $\lceil \cdot \rceil$ runden auf die nächste ganze Zahl bedeutet. Aus der T_{60} wird eine Abklingkonstante λ berechnet, mit welcher der exponentielle Abfall der Energie beschrieben werden kann. Wird der Lautsprecher zum Zeitpunkt $t_0 = 0$ abgeschaltet, klingt $P_{RR}(\mu, t)$ kontinuierlich über die Zeit t wie folgt ab

$$P_{RR}(\mu, t) = P_{RR}(\mu, t_0) \cdot e^{-\lambda t}. \quad (5.3)$$

Nach der Definition der Nachhallzeit ist bei $t = T_{60}$ die Energie auf -60 dB, d. h. den millionsten Teil des Anfangswerts abgefallen

$$P_{RR}(\mu, t = T_{60}) = e^{-\lambda T_{60}} \cdot P_{RR}(\mu, t_0) \quad (5.4)$$

$$1 \cdot 10^{-6} = e^{-\lambda T_{60}} \cdot 1, \quad (5.5)$$

woraus für λ folgt

$$\lambda = \frac{\ln(1 \cdot 10^{-6})}{T_{60}}. \quad (5.6)$$

Die Kopplungsfaktoren $A(\mu)$ werden mit der Welch-Methode berechnet [71]. Dazu wird die Impulsantwort in halb überlappende Blöcke der Länge N aufgeteilt. Die Blöcke werden mit einem Hamming-Fenster gewichtet und in den Frequenzbereich transformiert. Dort wird das Betragsquadrat gebildet, um das LDS zu erhalten. Anschließend wird der Mittelwert über die LDS der einzelnen Blöcke gebildet. Zur Berechnung der Kopplungsfaktoren wird das mittlere LDS mit einem nicht-kausalen Filter über die Frequenz geglättet. Im praktischen Einsatz hat sich diese Glättung als hilfreich erwiesen, da aufgrund von Messungenauigkeiten und Raumänderungen der gemessene Raum nicht exakt mit dem realen übereinstimmt.

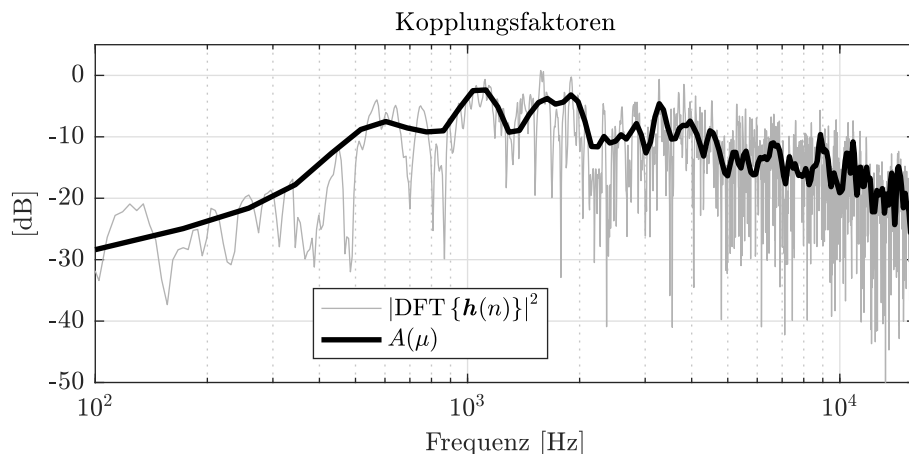


Abbildung 5.2: Die Kopplungsfaktoren von den hinteren Lautsprechern zum Fahrermikrofon des leeren Fahrzeugs.

Die so berechneten Kopplungsfaktoren sind in Abb. 5.2 der DFT der gemessenen Impulsantwort gegenübergestellt. Hier wurde die Impulsantwort des leeren Fahrzeugs von den hinteren Lautsprechern zum Fahrermikrofon aus Abb. 2.10a bzw. Abb. 2.10b auf S. 14 verwendet.

Mit den Parametern λ , K und $A(\mu)$ kann schließlich das LDS der Rückkopplung über folgenden rekursiven Zusammenhang geschätzt werden

$$\begin{aligned} P_{RR}(\mu, k) &\approx \hat{P}_{RR}(\mu, k) = b(\mu) \cdot \hat{P}_{XX}(\mu, k - K) + a \cdot \hat{P}_{RR}(\mu, k - 1) & (5.7) \\ a &= e^{\lambda \cdot \frac{1}{f_s}} \\ b(\mu) &= A(\mu). \end{aligned}$$

Hierbei wird das LDS des Lautsprechersignals $\hat{P}_{XX}(\mu, k)$ um K Blöcke verzögert, mit den Kopplungsfaktoren gewichtet und zum exponentiell abklingenden alten Filterausgang $\hat{P}_{RR}(\mu, k - 1)$ addiert. Das über Gl. (5.7) beschriebene Hallmodell kann also auch als IIR-Filter mit dem verzögerten Eingangssignal $\hat{P}_{XX}(\mu, k - K)$ und dem Ausgangssignal $\hat{P}_{RR}(\mu, k)$ verstanden werden. Die Einhüllende der Impulsantwort dieses Filters ist für ein Teilband schematisch in Abb. 5.3 dargestellt. Der auf diese Weise modellierte Übertragungspfad ist in Abb. 5.4 dem Spektrogramm der gemessenen Impulsantwort gegenüber gestellt. Hierbei wurden λ und K gemäß den Gleichungen (5.6) und (5.2) sowie der gemessenen Impulsantwort zu $\lambda \approx -114,9 \text{ s}^{-1}$ und $K = 1$ berechnet.

Das in Gl. (5.7) benötigte LDS des Lautsprechersignals $P_{XX}(\mu, k)$ wird durch IIR-Glättung des Kurzzeitleistungsdichtespektrums angenähert

$$P_{XX}(\mu, k) \approx \hat{P}_{XX}(\mu, k) = \overline{|X(\mu, k)|^2}, \quad (5.8)$$

ebenso gilt für das Fehlersignal

$$P_{EE}(\mu, k) \approx \hat{P}_{EE}(\mu, k) = \overline{|E(\mu, k)|^2}. \quad (5.9)$$

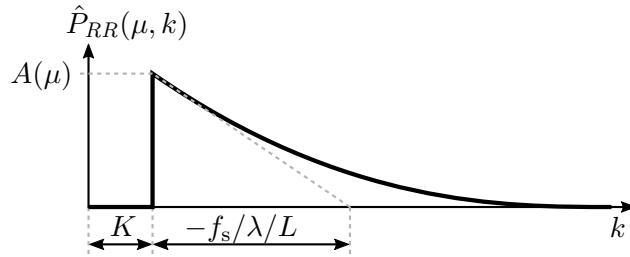


Abbildung 5.3: Schematische Darstellung des rekursiven Hallmodells und der zugehörigen Parameter.

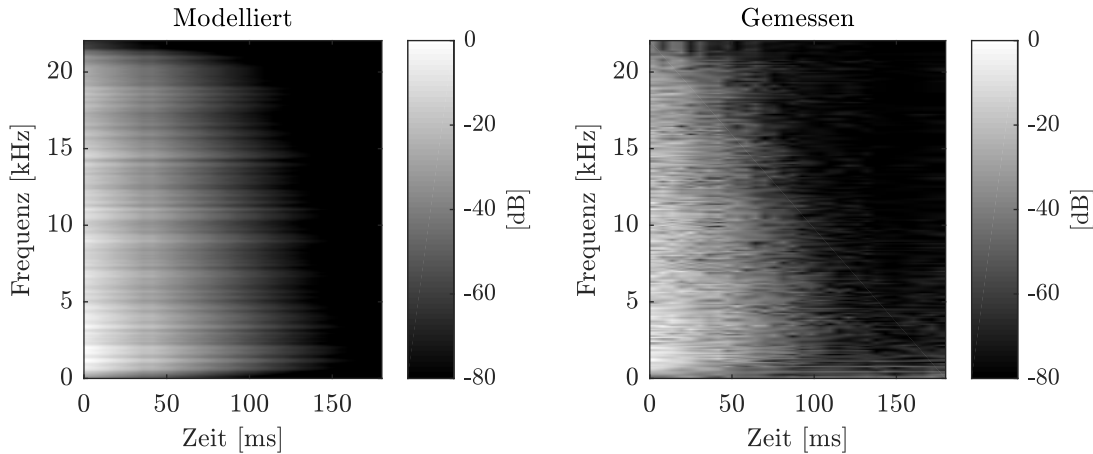


Abbildung 5.4: Spektrogramme des modellierten und des gemessenen Übertragungsverhaltens zwischen den hinteren Lautsprechern und dem Fahrermikrofon des leeren Fahrzeugs.

Damit kann schließlich das Wiener-Filter nach Gl. (5.1) berechnet werden. Das Filter wird zur klanglichen Optimierung um zwei Parameter ergänzt. Aufgrund fehlerhafter Schätzungen kann es vorkommen, dass das Filter in Sprachpausen in einzelnen Teilbändern fälschlicherweise zu wenig dämpft. Dies macht sich in Form von sog. tonalen Reststörungen bemerkbar. Dieser Effekt kann durch Überschätzung der geschätzten Rückkopplung mit dem Faktor c reduziert werden. Gl. (5.1) ändert sich damit zu

$$H_w(\mu, k) = 1 - c \cdot \frac{\hat{P}_{RR}(\mu, k)}{\hat{P}_{EE}(\mu, k)}. \quad (5.10)$$

Wird das Filter wie in Gl. (5.10) verwendet, werden wegen der möglichen sehr hohen Dämpfung des Filters auch Sprachanteile unterdrückt, was sich negativ auf das klangliche Ergebnis auswirkt. Aus diesem Grund wird die maximale Dämpfung des Filters auf den Wert $H_{FS, \min}$ begrenzt. Es folgt damit für das Wiener-Filter für die Rückkopplungsreduktion

$$H_{FS}(\mu, k) = \begin{cases} H_w(\mu, k) & \text{wenn } H_w(\mu, k) > H_{FS, \min}, \\ H_{FS, \min} & \text{sonst.} \end{cases} \quad (5.11)$$

Das Filter wird innerhalb der Filterbank für jeden Block neu berechnet. Die Änderungen des Filters zwischen zwei zeitlich aufeinander folgenden Blöcken können trotz des Fensters der *Overlap-Add*-Filterbank in Form von Kratzgeräuschen hörbar sein. Ist dies der Fall, können die Filterkoeffizienten zeitlich mit einem IIR-Filter geglättet werden

$$\overline{H_{\text{FS}}(\mu, k)} = \beta \cdot \overline{H_{\text{FS}}(\mu, k - 1)} + (1 - \beta) \cdot H_{\text{FS}}(\mu, k). \quad (5.12)$$

Für die Glättungskonstante haben sich hier Werte von $\beta < 0,5$ als passend herausgestellt. Größere Werte führen zu einer zu starken Verschmierung der Filterkoeffizienten. Die Maximaldämpfung des Filters wird idealerweise im Bereich $H_{\text{FS},\text{min}} = -20 \dots -10$ dB eingestellt, für die Überschätzung eignen sich Werte im Bereich $c = 1 \dots 3$.

5.2 Kombination von Rückkopplungskompensation und Rückkopplungsreduktion

Eine mögliche Kombination von modellbasierter Rückkopplungsreduktion und Rückkopplungskompensation wird in [74] umgesetzt. Dort wird das in Abschn. 5.1 beschriebene Verfahren parallel zu einem Rückkopplungskompensationsansatz betrieben. Bei genauer Betrachtung fällt allerdings auf, dass die Vorgehensweise nicht ideal ist, da hierbei Sprachanteile fälschlicherweise unterdrückt werden können. Der Grund ist, dass Rückkopplungsanteile sowohl phasenrichtig durch die Rückkopplungskompensation als auch spektral durch die Rückkopplungsreduktion vom Mikrofonsignal abgezogen werden. Vorausgesetzt, dass in beiden Fällen die Rückkopplung korrekt geschätzt wird, wird letztere also doppelt abgezogen. Um dies zu verhindern, wurden verschiedene Ansätze zur Modifikation der modellbasierten Rückkopplungsreduktion entwickelt, damit diese als Postfilter für die Rückkopplungskompensation eingesetzt werden kann.

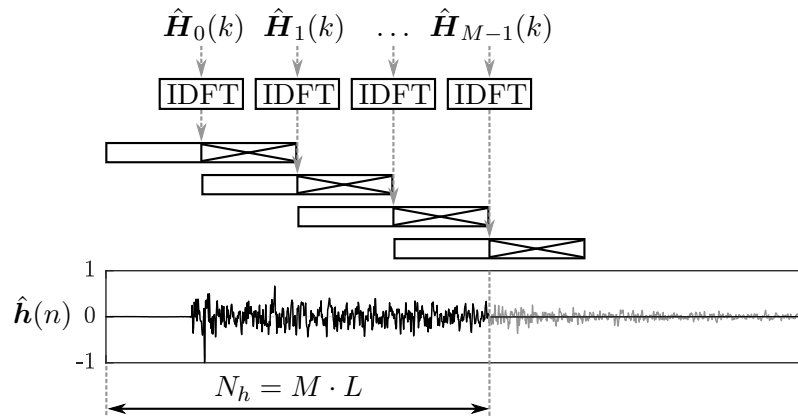


Abbildung 5.5: Zusammensetzen der geschätzten Impulsantwort aus den Teilfiltern des Rückkopplungskompensators. Der Kompensator deckt nur einen Teil der realen Impulsantwort (grau) ab.

Zur Veranschaulichung ist in Abb. 5.5 dargestellt, wie sich die geschätzte Impulsantwort (schwarz) im Zeitbereich aus den einzelnen Teilfiltern des Kompensators bei der *Overlap-Save*-Methode zusammensetzt. Wie bereits erwähnt, decken die M Teilfilter $\hat{\mathbf{H}}_0(k) \dots \hat{\mathbf{H}}_{M-1}(k)$ nur die ersten $N_h = M \cdot L$ Abtastwerte der realen Impulsantwort (grau) ab. Setzt man nun zunächst einen optimalen Abgleich des Kompensators voraus, d. h.

$$\hat{\mathbf{h}}(n) = \mathbf{h}(n), \quad (5.13)$$

so darf die Rückkopplungsreduktion in den ersten N_h Werten keine weiteren Rückkopplungsanteile mehr unterdrücken, da die Rückkopplung bereits vollständig kompensiert ist. Das in Gl. (5.7) bzw. Abb. 5.3 gezeigte Hallmodell muss daher in diesem Fall abgewandelt werden zu

$$\begin{aligned} \hat{P}_{RR}(\mu, k) &= b(\mu) \cdot \hat{P}_{XX}(\mu, k - M) + a \cdot \hat{P}_{RR}(\mu, k - 1) \\ a &= e^{\lambda \cdot \frac{L}{f_s}} \\ b(\mu) &= e^{\lambda \cdot \frac{L}{f_s} (M-K)} \cdot A(\mu), \end{aligned} \quad (5.14)$$

woraus das in Abb. 5.6 gezeigte Abklingverhalten folgt. Damit unterdrückt die Rückkopplungsunterdrückung ausschließlich den hinteren Teil der Impulsantwort, den der Kompensator aufgrund seiner begrenzten Länge nicht abbilden kann.

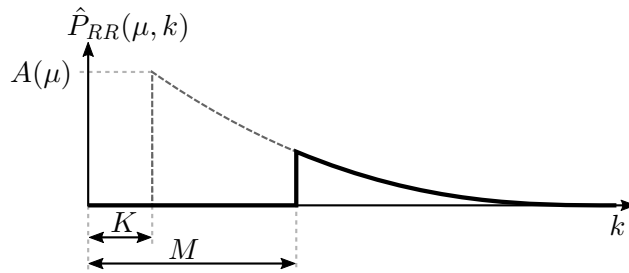


Abbildung 5.6: Das Abklingverhalten des rekursiven Hallmodells unter der idealen Bedingung, dass der Kompensator in den ersten M Blöcken perfekt abgleicht (schwarze Kurve).

5.3 Varianten des Hallmodells

Im realen Betrieb kann der Rückkopplungskompensator aufgrund der beschriebenen Korrelationsproblematik nicht perfekt abgleichen. Es bleibt immer ein Restfehler, welcher durch den Systemabstand $\|\mathbf{H}_{\Delta}(\mu, k)\|^2$ angegeben werden kann. Zusätzlich zu der im vorangegangenen Abschnitt beschriebenen Unterdrückung des späten Nachhalls ist es das Ziel des Postfilters, diesen Restfehler ebenfalls zu reduzieren. Insbesondere soll die Rückkopplungsreduktion immer dann aktiv sein, wenn der Rückkopplungskompensator schlecht abgeleitet ist. Umgekehrt soll die Rückkopplungsreduktion bei gutem Abgleich des Kompensators in den vorderen Teil der Impulsantwort nur wenig eingreifen.

5.3.1 Variante A

In Abschn. 4.4.1 wird erläutert, dass bei der Rückkopplungskompensation der Fehler in Form des Systemabstands gleichmäßig über die geschätzte Impulsantwort verteilt ist. Für eine korrekte Kombination von Rückkopplungsreduktion und -kompensation muss demnach der vordere Teil des Hallmodells dem Systemabstand entsprechen. Dies lässt sich durch folgende Rekursion realisieren

$$\hat{P}_{RR,1}(\mu, k) = \|\mathbf{H}_{\Delta}(\mu, k)\|^2 \cdot \sum_{i=0}^{M-1} \hat{P}_{XX}(\mu, k-i) \quad (5.15)$$

$$\hat{P}_{RR,2}(\mu, k) = b(\mu) \cdot \hat{P}_{XX}(\mu, k-M) + a \cdot \hat{P}_{RR,2}(\mu, k-1) \quad (5.16)$$

$$a = e^{\lambda \cdot \frac{T}{f_s}}$$

$$b(\mu) = e^{\lambda \cdot \frac{T}{f_s} (M-K)} \cdot A(\mu)$$

$$\hat{P}_{RR}(\mu, k) = \hat{P}_{RR,1}(\mu, k) + \hat{P}_{RR,2}(\mu, k). \quad (5.17)$$

Veranschaulichen lässt sich dies erneut mit der in Abb. 5.7 schematisch dargestellten Impulsantwort. Je besser der Kompensator abgleicht, desto kleiner wird der Systemabstand, womit sich das Hallmodell dem im vorangegangenen Abschnitt beschriebenen Modell nähert. Wie bereits beschrieben, ist der Systemabstand unbekannt. Da der Systemabstand jedoch auch für die in Kap. 4 beschriebene Schrittweitensteuerung des Rückkopplungskompensators benötigt wird, kann die dort berechnete Schätzgröße $\gamma(\mu, k)$ direkt auch für das hier vorgestellte Hallmodell übernommen werden.

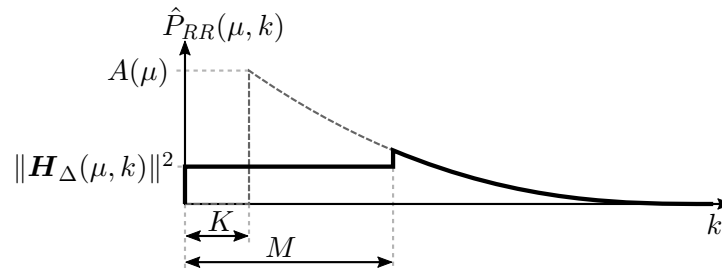


Abbildung 5.7: Variante A des Hallmodells, eingesetzt als Postfilter in Kombination mit der Rückkopplungskompensation.

5.3.2 Variante B

Bei einer zweiten Variante des Hallmodells wird der vordere Teil ebenfalls in Abhängigkeit des Systemabstands gesteuert. Allerdings wird hierbei nicht von einer gleichmäßigen Verteilung ausgegangen. Die Idee ist es, lediglich die Kopplungsfaktoren $A(\mu)$ um den Systemabstand zu dämpfen. Auch dies resultiert darin, dass der vordere Teil mit abnehmendem Systemabstand weniger stark in das Hallmodell eingeht. Allerdings bleibt das

exponentielle Abklingverhalten auch im vorderen Teil des Hallmodells bestehen. Ausgedrückt in rekursiver Form ergibt sich

$$\begin{aligned}\hat{P}_{RR}(\mu, k) &= b_1(\mu) \cdot \hat{P}_{XX}(\mu, k - K) + b_2(\mu) \cdot \hat{P}_{XX}(\mu, k - M) + a \cdot \hat{P}_{RR}(\mu, k - 1) \quad (5.18) \\ a &= e^{\lambda \cdot \frac{L}{f_s}} \\ b_1(\mu) &= A(\mu) \cdot \|\mathbf{H}_{\Delta}(\mu, k)\|^2 \\ b_2(\mu) &= e^{\lambda \cdot \frac{L}{f_s}(M-K)} \cdot (A(\mu) - b_1(\mu)),\end{aligned}$$

wobei der Faktor $b_1(\mu)$ die Dämpfung der Kopplungsfaktoren um den Systemabstand beschreibt. Ausgehend davon ergibt sich die in Abb. 5.8 dargestellte Einhüllende der Impulsantwort. Anders als bei der Variante A entspricht dieses Hallmodell für einen Systemabstand von $\|\mathbf{H}_{\Delta}(\mu, k)\|^2 = 0$ dB dem ursprünglichen Hallmodell aus Abschn. 5.1.

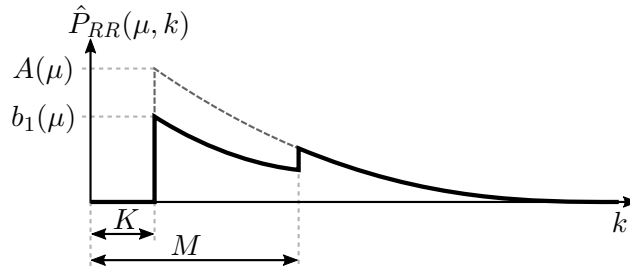


Abbildung 5.8: Variante B des Hallmodells. Die Kopplungsfaktoren werden um den Systemabstand gedämpft.

5.4 Ergebnisse und Simulationen

Zusammenfassend ist das schematische Blockdiagramm der modellbasierten Rückkopplungsreduktion in Abb. 5.9 gezeigt. Als Eingangsgrößen für die Schätzung des LDS der Restrückkopplung dienen der Systemabstand $\|\mathbf{H}_{\Delta}(\mu, k)\|^2$ und das geschätzte LDS des Lautsprechersignals $X(\mu, k)$. Das Wiener-Filter berechnet sich aus den geschätzten LDS von Restrückkopplung und Fehlersignal $E(\mu, k)$.

Die Performanz der verschiedenen Varianten des Hallmodells wird zunächst mittels objektiver Maße evaluiert. Um die erreichbare Qualität der verschiedenen Varianten zu vergleichen, wird in [24] zudem ein subjektiver Hörtest durchgeführt, dessen Ergebnisse am Ende dieses Abschnittes kurz präsentiert werden. In beiden Fällen dient die Aufnahme eines männlichen Sprechers, welcher sich auf dem Fahrersitz befindet, als lokales Sprachsignal. Um lediglich die Unterdrückung der Rückkopplung betrachten zu können, wird der Sprache kein Fahrgeräusch überlagert. Als reale Impulsantwort dient die bereits in Abschn. 5.1 beschriebene Impulsantwort des leeren Fahrzeugs. Die Parameter der Filterbänke und des Rückkopplungskompensators werden entsprechend Abschn. 3.5.2

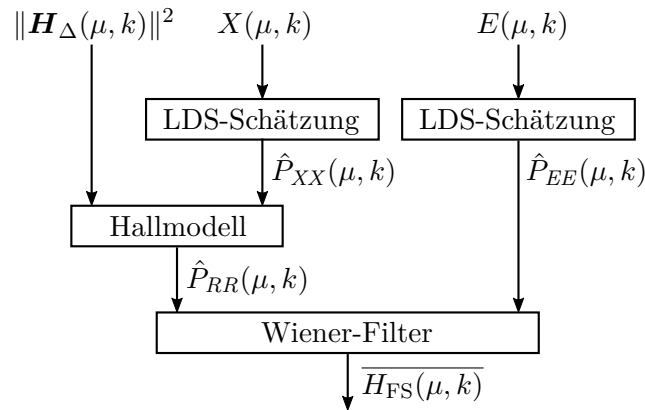


Abbildung 5.9: Blockdiagramm der für die modellbasierte Rückkopplungsunterdrückung benötigten Teilsysteme.

gewählt. Lediglich die Länge der vom Rückkopplungskompensator geschätzten Impulsantwort wurde auf $M = 4$ Teilfilter, bzw. $4L = 1024$ Abtastwerte gekürzt. Da damit der Kompensator jetzt nicht mehr den kompletten relevanten Bereich der Impulsantwort abdeckt, werden die Effekte des Postfilters deutlicher sichtbar. Die Stabilitätsgrenze ohne jegliche Algorithmen zur Rückkopplungsunterdrückung liegt bei einer Verstärkung von 0 dB. Darauf bezogen wird für die nachfolgenden Untersuchungen eine konstante Verstärkung von +2 dB eingestellt. Der Grund, weshalb hier keine höheren Verstärkungen erzielt werden können, ist die begrenzte Länge des Rückkopplungskompensators. Weil hier die Rückkopplung nicht vollständig kompensiert wird, befindet sich das System bereits bei +2 dB Verstärkung an der Stabilitätsgrenze. Für die nachfolgenden Untersuchungen spielt die absolute Verstärkung keine Rolle. Um den Effekt des Postfilters aufzuzeigen, ist es lediglich wichtig, dass Restrückkopplungen auftreten, was bei einem Betrieb nahe an der Stabilitätsgrenze gewährleistet ist. Es wird außerdem im Folgenden von einer erfolgreichen Konvergenz des Kompensators ausgegangen, d. h. der Systemabstand wird auf einem konstanten Wert gehalten. Raumänderungen werden nicht betrachtet.

5.4.1 Objektive Evaluierung

Bei der objektiven Evaluierung wurde untersucht, wie stark das Wiener-Filter die lokale Sprache angreift, verglichen mit der Unterdrückung der Restrückkopplung. Dazu werden die in der Simulation bekannten Signale $s(n) \circ \bullet S(\mu, k)$, d. h. die lokale Sprache, und $r(n) \circ \bullet R(\mu, k)$, d. h. die reale Rückkopplung, im Frequenzbereich mit den Koeffizienten des Wiener-Filters gewichtet. Anschließend werden sowohl die gefilterte als auch die ungefilterte lokale Sprache mit einer binären Sprachaktivitätsmaske $\Psi_s(\mu, k)$ gewichtet. Äquivalent dazu werden auch die gefilterte und die ungefilterte Rückkopplung mit einer binären Nachhallmaske $\Psi_r(\mu, k)$ gewichtet.

Berechnung der Masken

Die Maske $\Psi_s(\mu, k)$ ist zu Zeitpunkten detektierter Sprachaktivität eins, ansonsten null. Zur Berechnung der Maske wird in jedem Teilband der lokalen Sprache dem Minimum gefolgt, was in einer Schätzung $\hat{B}(\mu, k)$ des in der Sprache enthaltenen Restgeräusches resultiert, d. h.

$$\hat{B}(\mu, k) = f_{\min}(S(\mu, k)). \quad (5.19)$$

Immer dann, wenn $|S(\mu, k)|^2$ größer ist als $|\hat{B}(\mu, k)|^2 \cdot \beta_{\text{sad}}$, wird dies als Sprachaktivität interpretiert. β_{sad} ist hierbei eine Schwelle, mit der sich der Arbeitspunkt der Sprachaktivitätsdetektion einstellen lässt. Um zu verhindern, dass die Maske bei kurzfristigen Leistungseinbrüchen auf null springt, wird zudem eine Haltezeit eingefügt. Dadurch wird die Maske erst auf null gesetzt, wenn $|S(\mu, k)|^2$ für die Dauer mehrerer Zeitschritte unter $|\hat{B}(\mu, k)|^2 \cdot \beta_{\text{sad}}$ gefallen ist.

Genau umgekehrt dazu ist die Nachhallmaske $\Psi_r(\mu, k)$ immer während Sprachaktivität null und nur im Nachhall eins. Es besteht also zunächst der folgende Zusammenhang

$$\tilde{\Psi}_r(\mu, k) = 1 - \Psi_s(\mu, k) \quad (5.20)$$

zwischen den beiden Masken. Eine solche Implementierung würde allerdings dazu führen, dass die Nachhallmaske in langen Sprachpausen durchgehend aktiv wäre. Da jedoch explizit nur der Nachhall in die Maske eingehen soll, wird jedes Mal, wenn die Maske auf eins gesetzt wird, ein Zähler hochgezählt. Erreicht dieser Zähler die Nachhallzeit T_{60} , wird die Maske auf null gesetzt. Da in diesem Fall die Impulsantwort des leeren Fahrzeugs aus Abb. 2.10b auf S. 14 verwendet wird, bedeutet dies, dass $T_{60} \approx 120$ ms ist.

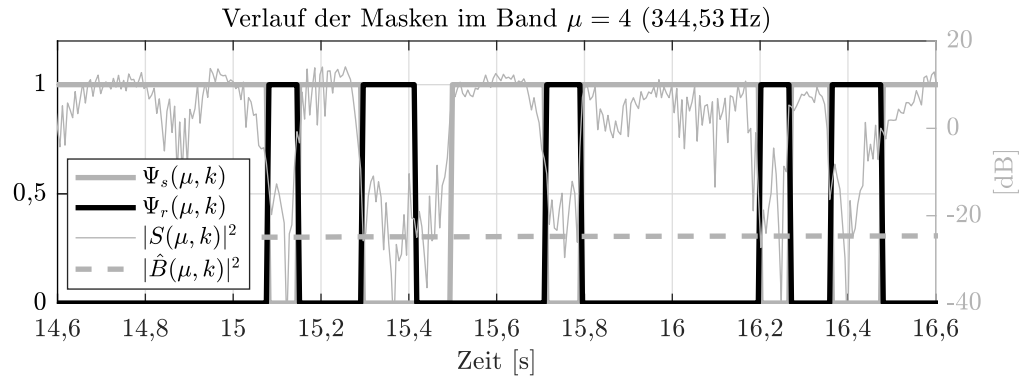
Beispielhaft sind in Abb. 5.10a die Verläufe der beiden Masken sowie die lokale Sprache im Teilband $\mu = 4$ (344,53 Hz) gezeigt. Die grau gestrichelte Linie markiert die Geräuschschätzung. Ab dem Zeitpunkt 15,3 s beginnt eine längere Sprachpause. Es ist zu erkennen, dass an dieser Stelle die Nachhallmaske $\Psi_r(\mu, k)$ für etwa 120 ms geöffnet bleibt, bevor sie bei etwa 15,42 s auf null fällt.

Der gleiche zeitliche Ausschnitt ist in Abb. 5.10b in Form von Spektrogrammen gezeigt. Dargestellt ist der sprachrelevante Frequenzbereich von etwa 300 Hz bis 3000 Hz. Auch hier zeigt sich, dass die Nachhallmaske $\Psi_r(\mu, k)$ der invertierten Sprachmaske $\Psi_s(\mu, k)$ entspricht, ausgenommen längerer Sprachpausen.

Maskierte Leistungsberechnung

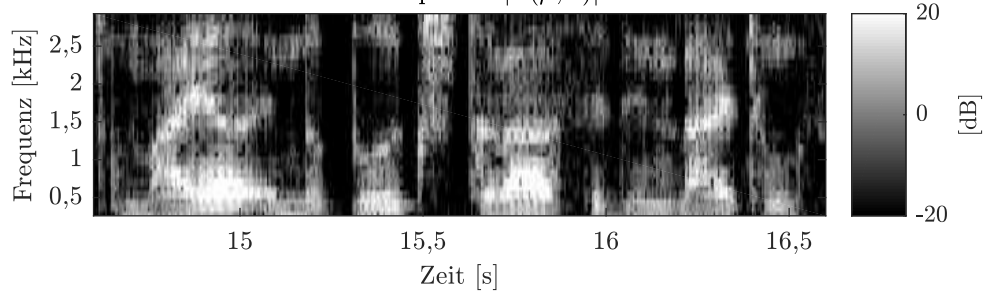
Wie eingangs beschrieben, werden schließlich die Sprache und Rückkopplung jeweils sowohl mit den Koeffizienten des Wiener-Filters $H_{\text{FS}}(\mu, k)$ als auch mit den Masken gewichtet. Dies ist schematisch im Blockdiagramm in Abb. 5.11 dargestellt. Nach der Gewichtung wird für alle Signale die mittlere Leistung im Frequenzband 300 Hz bis 3000 Hz berechnet. In Abb. 5.11 ist diese Mittelung in Zeit- und Frequenzrichtung als Erwartungswertbildung $E\{|\cdot|^2\}$ dargestellt.

Die Ergebnisse dieser Berechnungen sind in Tab. 5.1 dargestellt. Es werden drei verschiedene Abgleichzustände des Rückkopplungskompensators betrachtet. Dazu werden die drei verschiedenen Systemabstände $\|\mathbf{H}_{\Delta}(\mu, k)\|^2 = -12,8$ dB, $-22,8$ dB und $-32,8$ dB

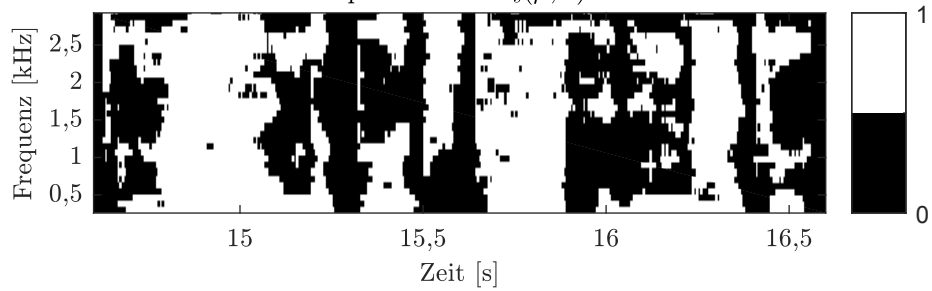


(a) Teilband.

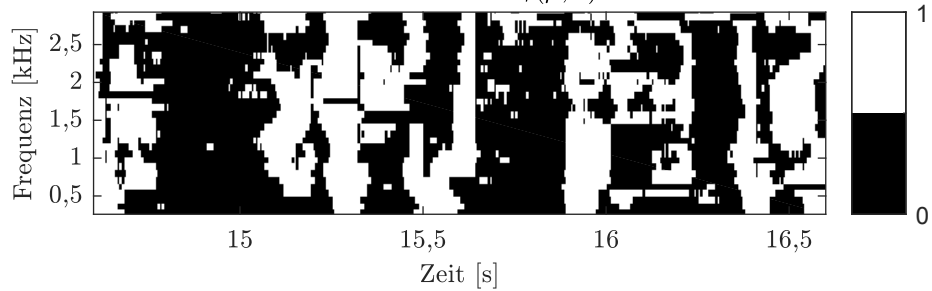
Lokale Sprache $|S(\mu, k)|$



Sprachmaske $\Psi_s(\mu, k)$



Nachhallmaske $\Psi_r(\mu, k)$



(b) Spektrogramme.

Abbildung 5.10: Ausschnitt des zeitlichen Verlaufs der Masken zur Evaluierung des Postfilters.

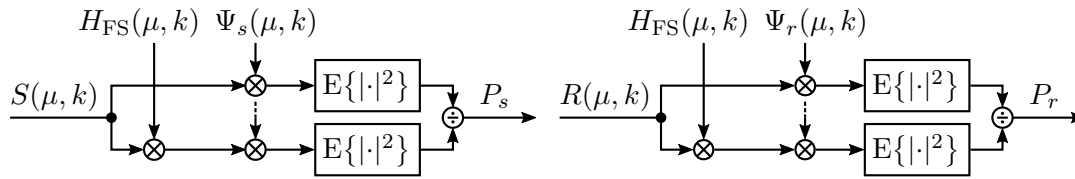


Abbildung 5.11: Schematisches Blockdiagramm zur Bestimmung der mittleren Leistungen.

jeweils fix eingestellt. Die Systemabstände werden lediglich über die Länge des Kompensators berechnet. Mit diesen Systemabständen werden die vier in den vorangegangenen Abschnitten vorgestellten Varianten des Hallmodells verglichen. *Original* bezeichnet dabei das in Abschn. 5.1 beschriebene, ursprüngliche Hallmodell, welches nicht für die Kombination mit einem Rückkopplungskompensator vorgesehen ist. *Ideal* bezeichnet das Hallmodell aus Abschn. 5.2, bei welchem davon ausgegangen wird, dass der Kompensator immer perfekt abgleicht. Die *Varianten A* und *B* sind schließlich die in Abschn. 5.3 vorgestellten Hallmodelle, welche speziell als Postfilter für die Rückkopplungskompensation entwickelt wurden.

$\ \mathbf{H}_\Delta(\mu, k)\ ^2 = -12,8 \text{ dB}$		P_s	P_r	P_r/P_s
Original	Gl. (5.7)	2,45 dB	4,76 dB	2,31 dB
Ideal	Gl. (5.14)	0,48 dB	4,49 dB	4,01 dB
Variante A	Gl. (5.17)	3,02 dB	9,32 dB	6,30 dB
Variante B	Gl. (5.18)	0,68 dB	4,58 dB	3,90 dB
$\ \mathbf{H}_\Delta(\mu, k)\ ^2 = -22,8 \text{ dB}$		P_s	P_r	P_r/P_s
Original	Gl. (5.7)	2,43 dB	8,08 dB	5,65 dB
Ideal	Gl. (5.14)	0,47 dB	7,33 dB	6,86 dB
Variante A	Gl. (5.17)	0,89 dB	7,92 dB	7,03 dB
Variante B	Gl. (5.18)	0,49 dB	7,35 dB	6,86 dB
$\ \mathbf{H}_\Delta(\mu, k)\ ^2 = -32,8 \text{ dB}$		P_s	P_r	P_r/P_s
Original	Gl. (5.7)	2,43 dB	8,23 dB	5,80 dB
Ideal	Gl. (5.14)	0,47 dB	7,49 dB	7,02 dB
Variante A	Gl. (5.17)	0,51 dB	7,55 dB	7,04 dB
Variante B	Gl. (5.18)	0,47 dB	7,49 dB	7,02 dB

Tabelle 5.1: Die über den Sprachbereich und die Zeit gemittelten Leistungen nach der Maskierung.

Zur Interpretation der Tabelle ist zunächst zu sagen, dass das Wiener-Filter idealerweise keine Sprachanteile unterdrücken soll. D. h. das Verhältnis von ungefilterter zu gefilterter Sprachleistung P_s soll bestenfalls 0 dB betragen. Auf der anderen Seite soll das Wiener-Filter möglichst viele Rückkopplungsanteile unterdrücken. Das bedeutet, dass das Verhältnis P_r von ungefilterter zu gefilterter Rückkopplungsleistung möglichst groß

sein soll. Das Verhältnis P_r/P_s wiederum gibt an, wie stark die Rückkopplung im Vergleich zur Sprache unterdrückt wird. Im schlechtesten Fall läuft dieses Verhältnis gegen 0 dB, was bedeutet, dass Sprache und Rückkopplung gleichermaßen angegriffen werden. Große Werte deuten hingegen an, wie viel mehr Rückkopplungsanteile als Sprachanteile unterdrückt werden.

Erwartungsgemäß schneidet das originale Hallmodell in allen Fällen am schlechtesten ab. Zwar wird Rückkopplung bei den Systemabständen -22,8 dB und -32,8 dB am stärksten unterdrückt (8,08 dB und 8,23 dB), allerdings wird auch die Sprache am stärksten angegriffen (2,43 dB), sodass das Verhältnis P_r/P_s in allen Fällen am schlechtesten ist, verglichen mit den anderen Modellen. Dass die Sprache mit diesem Hallmodell am stärksten angegriffen wird, zeigt auch der im nächsten Abschnitt beschriebene Hörtest. Es fällt bei diesem Hallmodell außerdem auf, dass P_r mit zunehmendem Systemabstand kleiner wird. Dies liegt wie beschrieben daran, dass bei einem geringen Systemabstand sowohl durch den Rückkopplungskompensator als auch durch die Rückkopplungsunterdrückung Rückkopplungsanteile abgezogen werden. Daher werden bei der Originalvariante fälschlicherweise insgesamt zu viel Rückkopplungsanteile abgezogen. Die vom Kompensator abgezogenen Anteile gehen jedoch mit in die Werte der Tabelle ein. Im Gegensatz dazu werden bei schlechtem Abgleich nur durch die Rückkopplungsunterdrückung Anteile abgezogen.

Die Unterschiede zwischen den verbleibenden drei Hallmodellen zeigen sich vor allem bei einem großen Systemabstand, da hier der vordere Teil des Hallmodells stärker in die Wiener-Filterung eingeht. Bemerkenswert ist, dass je größer der Systemabstand ist, desto stärker greift Variante A die Sprache an. Bei $\|\mathbf{H}_\Delta(\mu, k)\|^2 = -12,8$ dB wird die Sprache mit 3,02 dB sogar stärker gedämpft als mit dem originalen Hallmodell (2,45 dB). Das hat auch zur Folge, dass die wahrgenommene Verstärkung insgesamt zurück geht, da das komplette Signal stark gedämpft wird. Der Grund für die starke Dämpfung mit Variante A ist, dass bei einem großen Systemabstand der vordere Teil hier stärker in das Modell eingeht. Die Variante *Ideal* und die Variante B liegen meist dicht beieinander. Speziell beim kleinsten Systemabstand $\|\mathbf{H}_\Delta(\mu, k)\|^2 = -32,8$ dB ist kein Unterschied feststellbar. Nur bei großem Systemabstand laufen die beiden Varianten geringfügig auseinander, hier ist das Verhältnis P_r/P_s bei Variante B um etwa 0,11 dB schlechter.

5.4.2 Subjektive Evaluierung

Im Rahmen der Veröffentlichung [24] wird ein Hörtest zum Vergleich der verschiedenen Hallmodelle durchgeführt. Wie auch bei der im vorangegangenen Abschnitt beschriebenen objektiven Evaluierung, ist es das Ziel dieses Tests herauszufinden, wie stark die Sprache von den verschiedenen Hallmodellen angegriffen wird und ob Restrückkopplungsanteile hörbar sind. Für den Hörtest wird das Szenario mit dem geringsten Systemabstand $\|\mathbf{H}_\Delta(\mu, k)\|^2 = -32,8$ dB verwendet.

Der Hörtest wird als *Mean Opinion Score* (MOS) Test aufgesetzt [33]. Insgesamt nahmen an dem Hörtest 26 normalhörende Probanden teil. Der Hörtest wird mit Kopfhörern durchgeführt. Den Probanden wird zunächst die unverarbeitete lokale Sprache als optimales Referenzsignal vorgespielt. Anschließend folgt das Postfilter mit den vier

verschiedenen Hallmodellen in Kombination mit dem Rückkopplungskompensator. Zusätzlich wird nur der Kompensator ohne Postfilter ebenfalls miteinbezogen. Diese fünf Aufnahmen werden in zufälliger Reihenfolge vorgespielt. Ein solcher Block, bestehend aus insgesamt sechs Aufnahmen (Referenz und fünf Varianten), wird jeweils mit drei weiblichen und zwei männlichen Sprechern präsentiert. Der Sprecher jeder Aufnahme sagt jeweils zwei kurze, zusammenhangslose Sätze. Der erste Durchgang wird mit einem weiblichen Sprecher durchgeführt und fließt nicht in die Auswertung mit ein. Er dient dazu, den Probanden die Spanne der verschiedenen Qualitätsstufen aufzuzeigen. Diese können mit folgender fünfstufigen MOS Skala bewertet werden:

5. Ausgezeichnet - Die Sprache klingt wie das unverarbeitete Signal.
4. Gut - Die Sprache ist leicht angegriffen aber klingt noch natürlich.
3. Ordentlich - Die Sprache ist angegriffen, aber die Artefakte stören nicht.
2. Mäßig - Die Sprachqualität nimmt weiter ab, störende Artefakte sind deutlich wahrnehmbar.
1. Mangelhaft - Die Sprache ist stark angegriffen.

Das Ergebnis der Auswertung ist in Abb. 5.12 als Boxplot dargestellt. In diesem markieren die rechteckigen Boxen den Bereich, innerhalb welchem die mittleren 50% der Messdaten liegen. Die 25% der größten bzw. kleinsten Daten liegen jeweils oberhalb bzw. unterhalb der Box und werden durch die gestrichelt dargestellten Antennen repräsentiert. Die schwarze, dicke Linie innerhalb der Box markiert den Median. In der Abb. 5.12 sind die Ergebnisse aller fünf untersuchten Varianten gezeigt.

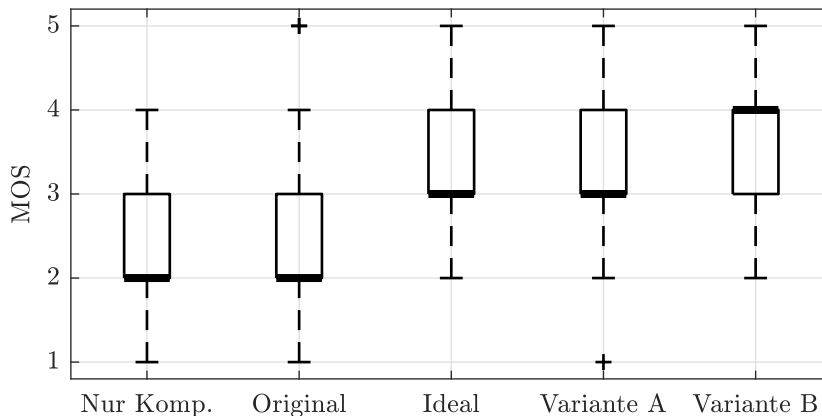


Abbildung 5.12: Ergebnis des Hörtests zur Evaluierung der Postfiltervarianten mit 26 normalhörenden Probanden. Der Kompensator alleine („Nur Komp.“) ist dem Postfilter mit dem originalen Hallmodell sowie den drei abgewandelten Hallmodellen gegenübergestellt.

Zunächst ist zu erkennen, dass der Kompensator ohne Postfilter und der Kompensator mit Postfilter mit originalem Hallmodell nahe beieinander liegen. Der Median ist in

beiden Fällen zwei. Dass diese beiden Varianten eng beieinander liegen, liegt daran, dass beim reinen Kompensator noch Restrückkopplungen zu hören sind. Beim Postfilter mit originalem Hallmodell sind die Restrückkopplungen zwar unterdrückt, allerdings ist auch die Sprache stark angegriffen. In beiden Fällen führt dies dazu, dass die Sprachqualität schlechter bewertet wird.

Man erkennt weiter, dass die beiden Hallmodell-Varianten A und B sowie das Hallmodell, welches von einem idealen Kompensatorabgleich ausgeht, ebenfalls nahe beieinander liegen. Der einzige Unterschied ist, dass der Median bei Variante B bei vier liegt, während er bei den anderen beiden Modellen drei beträgt. Der Grund ist hier der geringe Systemabstand von $\|\mathbf{H}_{\Delta}(\mu, k)\|^2 = -32,8 \text{ dB}$. Dieser führt dazu, dass sich die drei Hallmodelle klanglich nur unwesentlich unterscheiden. Wie die objektive Evaluierung sollte man also auch den Hörtest idealerweise mit verschiedenen Systemabständen durchführen.

Deutlicher ist der Unterschied zwischen den speziell für das Postfilter entwickelten Hallmodellen zum originalen Hallmodell, bzw. zum Kompensator ohne Postfilter. Hier ist zu erkennen, dass sowohl die Boxen als auch die Antennen um eine MOS-Stufe nach oben verschoben sind. Das bedeutet, dass die Sprachqualität hier besser bewertet wird.

5.5 Zusammenfassung

Ausgehend von dem bekannten Ansatz der modellbasierten Rückkopplungsreduktion nach [78] wurde ein Postfilter für die Rückkopplungskompensation entwickelt. Dazu wurde das Verfahren so modifiziert, dass das Hallmodell abhängig vom Abgleichzustand der Rückkopplungskompensation gesteuert wird. Hierfür wurden zwei verschiedene Varianten des Hallmodells entwickelt. Mit beiden Varianten können die Restrückkopplungen deutlich reduziert werden, sodass sich sowohl objektiv als auch subjektiv die Sprachqualität verbessert. Insgesamt kann festgestellt werden, dass der Hörtest das Ergebnis der objektiven Evaluierung bestätigt. Da die beiden Varianten A und B zu ähnlichen Ergebnissen führen, kann keine pauschale Empfehlung gegeben werden, welche Variante zu bevorzugen ist. Für Demonstrationen, die im Rahmen der vorliegenden Arbeit im Versuchsfahrzeug durchgeführt worden sind, ist die Variante A umgesetzt worden, da diese die mathematisch korrekte Lösung darstellt.

Kapitel 6

Verfahren zur Verbesserung der Sprachqualität

Während in den vorangegangenen Kapiteln der Fokus auf der Kompensation bzw. Reduktion der Rückkopplung liegt, werden in diesem Kapitel zwei Verfahren vorgestellt, mit denen der Klangeindruck des Systems beim Zuhörer verbessert werden kann. In Abschn. 6.1 wird ein Verfahren beschrieben, mit welchem das Lautsprechersignal entzerrt wird. Das Ziel des in Abschn. 6.2 beschriebenen Verfahrens ist es, die Sprachverständlichkeit in geräuschbehafteter Umgebung zu verbessern.

6.1 Automatische Entzerrung

Wie bereits in Abschn. 3.1 erläutert, schwingt sich eine geschlossene elektro-akustische Schleife vor allem bei Frequenzen auf, an denen der Betragsfrequenzgang des Rückkopplungspfad Überhöhungen aufweist. Aus diesem Grund werden bei der Entzerrung existierender ICC-Systeme an den Frequenzen dieser Überhöhungen Kerbfilter auf das Lautsprechersignal angewandt. Auf diese Weise werden die Resonanzfrequenzen des Raumes gedämpft und die Stabilität des Systems kann in geringem Maße verbessert werden. Klassischerweise werden diese Kerbfilter einmalig anhand vorab gemessener Impulsantworten parametrisiert. Dieses Vorgehen hat jedoch zwei Nachteile. Zum einen wird lediglich das Übertragungsverhalten zwischen den Lautsprechern und dem Sprechermikrofon entzerrt. Wünschenswert im Sinne eines optimalen Klangerlebnisses des Zuhörers ist hingegen eine Entzerrung auf das Zuhörerohr. Dies wird durch eine reine Entzerrung des Rückkopplungspfad nicht zwangsläufig erreicht. Zum anderen ändern sich durch Raumänderungen auch die Resonanzfrequenzen, weshalb ein automatisches Verfahren, welches sich zur Laufzeit selbst parametrisiert, für die Entzerrung zu bevorzugen ist. Dies kann beispielsweise mittels Aufschwingdetektoren realisiert werden [69]. Bei diesen Verfahren wird ein Kerbfilter adaptiv bei Frequenzen platziert, an denen zuvor ein Aufschwingen detektiert wurde.

Für Freisprecheinrichtungen existieren Verfahren, welche die beiden genannten Nachteile adressieren. So wird z. B. in [30, 56] ein Verfahren patentiert, welches eine Entzerrung auf Basis des Echokompensationsfilters vornimmt. Eine Entzerrung auf das Zuhörerohr kann mit diesem Verfahren jedoch nur näherungsweise stattfinden, nämlich unter der Annahme, dass sich das Mikrofon des lokalen Teilnehmers in der Nähe seiner Ohren befindet.

Mit der in Kap. 4 beschriebenen Rückkopplungskompensation ist das ICC-System auch ohne eine Entzerrung des Rückkopplungspfad bis hin zu großen Verstärkungen stabil. Dies macht eine Entzerrung des Rückkopplungspfad im Sinne der Stabilität überflüssig.

Aus diesem Grund liegt der Fokus in diesem Kapitel auf einem Verfahren zur automatischen Entzerrung des Frequenzgangs am Zuhörerohr. Das Ziel dieses Verfahrens ist es, einen möglichst natürlichen Klangeindruck am Zuhörerplatz zu erreichen. Im Gegensatz zu den oben erwähnten Patentschriften kann hier jedoch nicht das Rückkopplungskompensationsfilter für den Filterentwurf des Entzerrers herangezogen werden. Stattdessen wird zweites adaptives Filter benötigt, welches den Pfad zwischen Lautsprecher und Zuhörerohr schätzt. Dieses nachfolgend vorgestellte Verfahren wurde auch in [9] veröffentlicht.

6.1.1 Adaptive Filter zum Schätzen des kurzen Pfades

Bei einem Fahrzeug mit einem ICC-System überlagern sich am Ohr des Zuhörers der Direktschall des lokalen Sprachsignals sowie das Lautsprechersignal, gefaltet mit der Impulsantwort zwischen Lautsprecher und Ohr. Vor allem in großen Fahrzeugen, wie dem in Kap. 2 vorgestellten Mercedes-Benz Viano, zeigen Messungen, dass die lokale Sprache des Fahrers beim Zuhörer der dritten Sitzreihe um etwa 20 dB bis 30 dB gedämpft ankommt. Das Lautsprechersignal am Zuhörerohr ist zum einen gefärbt durch den Frequenzgang der Lautsprecher selbst und zum anderen durch die akustischen Eigenschaften des Fahrzeuginnenraums. Die Überlagerung all dieser Effekte führt dazu, dass sich die Sprache beim Zuhörer oftmals unnatürlich anhört.

Ziel des hier vorgestellten Verfahrens ist es, das Lautsprechersignal so zu entzerren, dass sich beim Zuhörer ein vorgegebener Zielfrequenzgang einstellt. Im einfachsten Fall kann dieser Zielfrequenzgang beispielsweise weiß sein. Um das Entzerrfilter entsprechend entwerfen zu können, muss zunächst das Übertragungsverhalten zwischen Lautsprecher und Zuhörerohr mit einem adaptiven Filter geschätzt werden. Für dieses adaptive Filter muss allerdings das Signal am Ohr des Zuhörers bekannt sein. Da dies mangels Sensoren im Ohr nicht direkt möglich ist, wird stattdessen angenommen, dass mit einem Mikrofon im Dachhimmel über dem Zuhörer eine gute Annäherung erreicht wird (vgl. auch Abb. 2.2 auf S. 7). Es wird angenommen, dass die Impulsantwort von den Lautsprechern zu diesem Mikrofon näherungsweise der Impulsantwort von den Lautsprechern zu den Ohren entspricht. Eine zweite notwendige Annahme ist, dass der Frequenzgang des Mikrofons im Sprachbereich linear ist. Da die hier verwendeten Freisprechemikrofone im Bereich von 100 Hz bis 10 kHz einen leichten Anstieg zeigen, gilt dies nur in grober Näherung.

Das adaptive Filter für die automatische Entzerrung ist zusammen mit dem adaptiven Filter für die Rückkopplungskompensation in Abb. 6.1 dargestellt. In der Abbildung bezeichnet $\hat{h}_k(n)$ die geschätzte Impulsantwort des kurzen Pfades. Die Signale $y_k(n)$ und $e_k(n)$ bezeichnen entsprechend das Mikrofonsignal am Zuhörerplatz und das Fehlersignal des adaptiven Filters. Erneut sind die Verstärkung und die Verzögerung aufgrund der Blockverarbeitung sowie die evtl. zusätzlich benötigten Signalverarbeitungsmodulare in der Impulsantwort des Vorwärtspfades $h_v(n)$ zusammengefasst. Die Impulsantwort des Entzerrers ist $h_{eq}(n)$. Der grau gestrichelte Pfeil deutet an, dass der Entzerrer durch die Schätzung des kurzen Pfades gesteuert wird. Am Lautsprecher wird schließlich das entzerrte Signal $x_{eq}(n)$ ausgegeben.

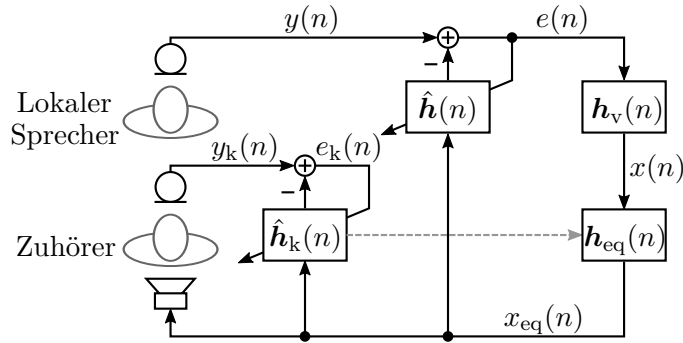


Abbildung 6.1: Blockdiagramm eines ICC-Systems mit automatischer Entzerrung des Zuhörerpfades. Für eine übersichtlichere Darstellung sind alle Signale im Zeitbereich dargestellt.

Die Steuerung des adaptiven Filters $\hat{h}_k(n)$ gestaltet sich als weniger kompliziert als die des Filters für die Rückkopplungskompensation. Der Grund ist die eingangs erwähnte Dämpfung der lokalen Sprache am Zuhörmikrofon von 20 dB bis 30 dB. Diese sorgt dafür, dass das Filter ungestört konvergieren kann. Außerdem ist dadurch auch die Korrelation zwischen Anregungs- bzw. Lautsprechersignal und lokaler Sprache geringer.

Zur Verringerung der Rechenleistung wird das adaptive Filter, wie in Abschn. 3.3.2 beschrieben, mittels einer *Overlap-Save*-Filterbank im Frequenzbereich implementiert. Das entsprechende Blockdiagramm ist in Abb. 6.2 gezeigt. In der Abbildung ist $h_k(n)$ die zu schätzende Impulsantwort des kurzen Pfades und $s_{y_k}(n)$ der Anteil der lokalen Sprache, welcher am Zuhörmikrofon ankommt. Zur Verbesserung der Konvergenz des

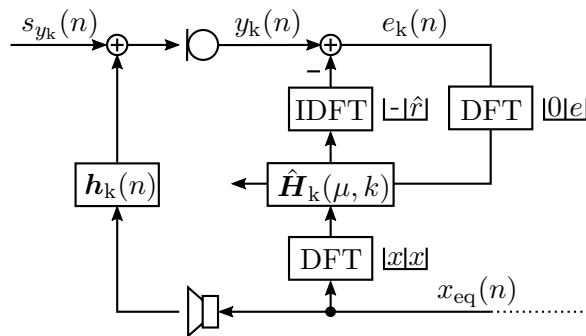


Abbildung 6.2: Schematische Darstellung der Frequenzbereichsimplementierung des adaptiven Filters zur Schätzung des kurzen Pfades.

adaptiven Filters wird die Schrittweite des Filters $\alpha_k(\mu, k)$ mit der Approximation der optimalen Schrittweite nach Gl. (4.14) gesteuert. Der zur Berechnung der Schrittweite benötigte Systemabstand wird mit dem in Abschn. 4.4.2 beschriebenen Verfahren ge-

schätzt. Mathematisch ausgedrückt bedeutet dies

$$\alpha_k(\mu, k) = \alpha_{\text{fix}} \cdot \frac{|X_{\text{eq}}(\mu, k)|^2}{|E_k(\mu, k)|^2} \cdot \gamma_k(\mu, k), \quad (6.1)$$

wobei für den geschätzten Systemabstand gilt

$$\gamma_k(\mu, k) \approx \overline{\|\mathbf{H}_k(\mu, k) - \hat{\mathbf{H}}_k(\mu, k)\|^2}. \quad (6.2)$$

Wie bereits bei der Rückkopplungskompensation, wird die fixe Schrittweite auf den Wert $\alpha_{\text{fix}} = 0,3$ eingestellt.

Der vom adaptiven Filter zu schätzende Pfad wird durch die in Abb. 2.9b auf S. 13 dargestellten Impulsantworten von den hinteren Lautsprechern zu den hinteren Mikrofonen beschrieben. Um den relevanten Teil der Impulsantwort abzubilden, werden auch hier $M = 8$ Teilfilter verwendet. Die weiteren Filterbankparameter entsprechen ebenfalls den bereits in Tab. 3.2 gezeigten.

6.1.2 Zielfrequenzgang des Entzerrfilters

Als Zielfrequenzgang soll der Frequenzgang des adaptiven Filters $\hat{\mathbf{h}}_k(\mu, k)$ so entzerrt werden, dass das Produkt der Betragsfrequenzgänge von Entzerrfilter und adaptivem Filter näherungsweise weiß wird. Um dies zu erreichen, muss das Entzerrfilter invers zum adaptiven Filter sein. Der Betragsfrequenzgang des adaptiven Filters nach erfolgter

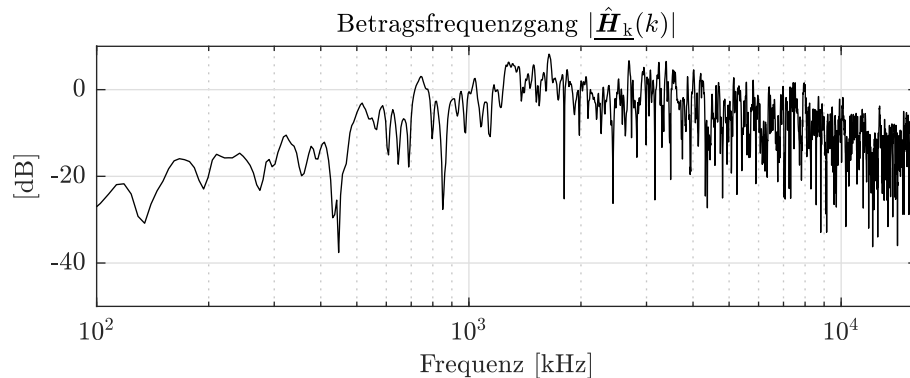


Abbildung 6.3: Der Betragsfrequenzgang des konvergierten adaptiven Filters nach ca. 19,5 s Simulationszeit.

Konvergenz ist in Abb. 6.3 dargestellt. Das adaptive Filter setzt sich aus den M Teilfiltern $\hat{\mathbf{H}}_{k,m}(k)$ zusammen. Wegen der verwendeten Filterbank hat jedes dieser Teilfilter die DFT-Auflösung $N = 512$, bei halber Überlappung im Zeitbereich. Für das Entzerrfilter wird allerdings der Frequenzgang über die gesamte Länge der Impulsantwort benötigt. Der Grund hierfür ist, dass auch die späten Reflexionen Einfluss auf die klangliche Färbung haben können. Um den Frequenzgang des adaptiven Filters in hoher Auflösung über die komplette Impulsantwort zu erhalten, muss letztere, wie bereits in Abb. 5.5 auf

S. 86 gezeigt, zunächst aus den Teilfiltern zusammengesetzt werden. Anschließend wird die Zeitbereichsimpulsantwort $\hat{\mathbf{h}}_k(n)$ mittels DFT der Länge $N_h = 2048$ zurück in den Frequenzbereich transformiert

$$\underline{\hat{\mathbf{H}}}_k(k) = \left[\underline{\hat{H}}_k(0, k), \underline{\hat{H}}_k(1, k), \dots, \underline{\hat{H}}_k(N_h - 1, k) \right]^T = \text{DFT} \left\{ \hat{\mathbf{h}}_k(n) \right\}. \quad (6.3)$$

Zur besseren Unterscheidung werden im Folgenden Frequenzbereichsvariablen mit großer DFT-Auflösung ($N_h = 2048$) mit einem Unterstrich ($\underline{\cdot}$) gekennzeichnet.

Der erste Schritt zur Berechnung des Entzerrfilters ist es, den Betrag von $\underline{\hat{\mathbf{H}}}_k(k)$ auf den Maximalwert 1 bzw. 0 dB zu normieren

$$\underline{H}_{\text{norm}}(\mu, k) = \frac{|\underline{\hat{H}}_k(\mu, k)|}{\max \left\{ |\underline{\hat{\mathbf{H}}}_k(k)| \right\}}. \quad (6.4)$$

Hierbei ist mit $\max \left\{ |\underline{\hat{\mathbf{H}}}_k(k)| \right\}$ das größte Element des Vektors $|\underline{\hat{\mathbf{H}}}_k(k)|$ gemeint. Anschließend wird $\underline{H}_{\text{norm}}(\mu, k)$ auf einen Minimalwert begrenzt. Dies ist notwendig, da aufgrund der nachfolgenden Invertierung an den Stellen, an denen der Frequenzgang des adaptiven Filters Kerben aufweist, das Entzerrfilter Spitzen hat. Liegt eine solche Spitze bei einer Resonanzfrequenz des Rückkopplungspfades, kann dies zur Folge haben, dass die MSV bei dieser Frequenz überschritten und das System instabil wird. Damit die Höhe dieser Spitzen begrenzt werden kann, wird der Parameter d eingeführt

$$\underline{H}_{\text{lim}}(\mu, k) = \begin{cases} \underline{H}_{\text{norm}}(\mu, k) & \text{wenn } \underline{H}_{\text{norm}}(\mu, k) > d, \\ d & \text{sonst,} \end{cases} \quad (6.5)$$

wobei sich in der Praxis ein Wert von $d = -20$ dB als praktikabel herausgestellt hat. Im nächsten Schritt wird $\underline{H}_{\text{lim}}(\mu, k)$ invertiert

$$\underline{H}_{\text{inv}}(\mu, k) = \frac{1}{\underline{H}_{\text{lim}}(\mu, k)} \quad (6.6)$$

und auf den Mittelwert im Intervall $\mu_s = [\mu_u, \mu_o]$ normiert

$$\underline{H}_{\text{inv, norm}}(\mu, k) = \frac{\underline{H}_{\text{inv}}(\mu, k)}{\frac{1}{N_{\mu_s}} \sum_{\mu=\mu_u}^{\mu_o} \underline{H}_{\text{inv}}(\mu, k)}. \quad (6.7)$$

Das Intervall μ_s entspricht dem Sprachbereich bzw. dem Bereich, in welchem der Frequenzgang des Mikrofons näherungsweise linear ist. In dieser Arbeit wird ein Bereich von $\mu_u = 5$ (ca. 100 Hz) bis $\mu_o = 464$ (ca. 10.000 Hz) verwendet. $N_{\mu_s} = \mu_o - \mu_u + 1$ ist die Anzahl diskreter Frequenzwerte im Intervall μ_s . Werte unterhalb bzw. oberhalb dieses Intervalls werden auf den Wert eins gesetzt

$$\underline{H}_{\text{eq}}(\mu, k) = \begin{cases} \underline{H}_{\text{inv, norm}}(\mu, k) & \text{wenn } \mu_u \leq \mu \leq \mu_o \\ 1,0 & \text{sonst.} \end{cases} \quad (6.8)$$

Hierdurch werden die außerhalb des Intervalls liegenden Frequenzbereiche nicht entzerrt. Auf diese Weise wird verhindert, dass Frequenzbereiche, die aufgrund des begrenzten Mikrofonfrequenzgangs gedämpft sind, fälschlicherweise verstärkt werden. Um die Sprungstellen an den Intervallgrenzen zu glätten und um enge Spitzen bzw. Kerben abzuflachen, wird $\overline{H_{\text{eq}}}(\mu, k)$ mit einem nicht-kausalen Filter entlang der Frequenzachse geglättet, was in $\overline{\overline{H_{\text{eq}}}}(\mu, k)$ resultiert.

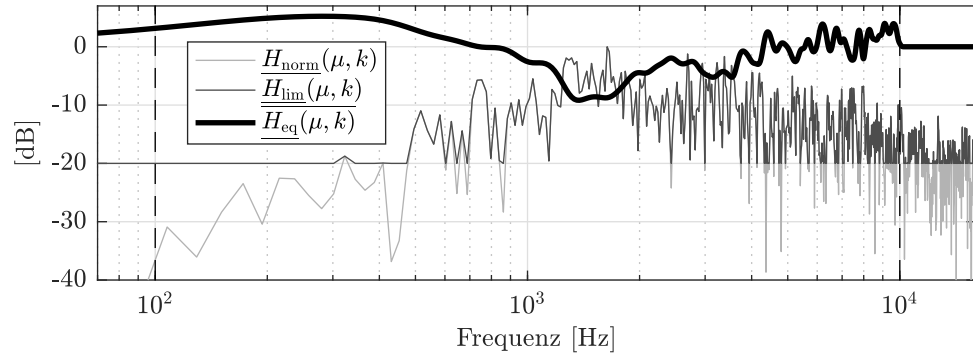


Abbildung 6.4: Betragsfrequenzgänge zur Berechnung des Entzerrfilters.

Die Frequenzgänge von $\overline{H_{\text{norm}}}(\mu, k)$, $\overline{H_{\text{lim}}}(\mu, k)$ und $\overline{H_{\text{eq}}}(\mu, k)$ sind für das in Abb. 6.3 gezeigte adaptive Filter in Abb. 6.4 dargestellt.

Unterabtastung von $\overline{H_{\text{eq}}}(\mu, k)$ um den Faktor $M/2$ entlang der diskreten Frequenz führt schließlich auf das Filter $\overline{\overline{H_{\text{eq}}}}(\mu, k)$. Dieses wird innerhalb der Filterbank des Vorwärtspfads multiplikativ auf das Lautsprechersignal $X(\mu, k)$ angewandt, sodass gilt

$$X_{\text{eq}}(\mu, k) = \overline{\overline{H_{\text{eq}}}}(\mu, k) \cdot X(\mu, k). \quad (6.9)$$

6.1.3 Ergebnis

In der Simulation werden die akustischen Pfade mit den in Kap. 2 gemessenen Impulsantworten modelliert. Der lokale Sprecher befindet sich auf dem Fahrersitz, der Rückkopplungspfad ist entsprechend der Pfad von den hinteren Lautsprechern zum Fahrer-mikrofon. Der Zuhörer sitzt auf dem linken Platz der dritten Sitzreihe, folglich wird auf diese Hörposition entzerrt. Der lokalen Sprache ist Hintergrundgeräusch überlagert, aufgenommen bei 100 km/h Autobahnfahrt. Mit diesem Geräusch stellt sich am Fahrer-mikrofon ein SNR von etwa 2,3 dBA ein, am Mikrofon hinten links über dem Zuhörerplatz beträgt das SNR ca. -15,2 dBA. Die Verstärkung des ICC-Systems wird auf 10 dB eingestellt, bezogen auf eine Stabilitätsgrenze von 0 dB wenn keine weiteren Maßnahmen zur Unterdrückung der Rückkopplung zum Einsatz kommen.

In Abb. 6.5 ist das zeitliche Konvergenzverhalten des adaptiven Filters in Form des über alle Frequenzen gemittelten Systemabstands dargestellt. Es ist zu erkennen, dass das Filter bereits nach 1 s auf -20 dB abgeglichen hat (graue gestrichelte Linie). Dieser schnelle Abgleich ist wichtig, da das Entzerrfilter bei einem schlecht abgeglichenen ad-

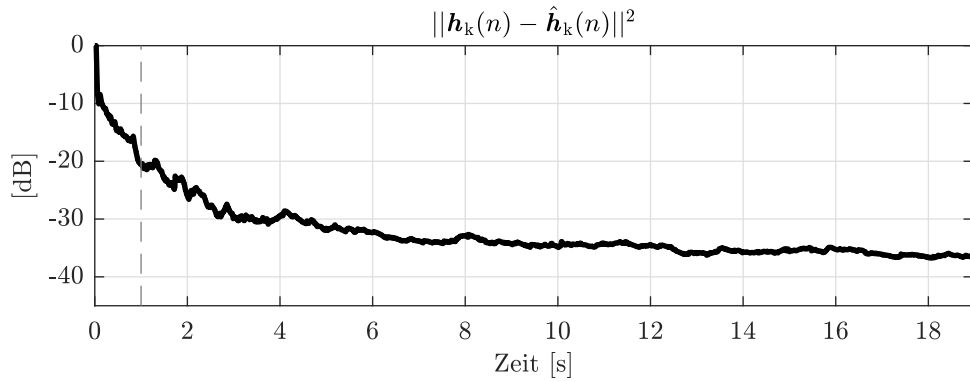


Abbildung 6.5: Das Konvergenzverhalten des adaptiven Filters für die automatische Entzerrung.

aptiven Filter den Pfad nicht korrekt entzerrt und damit das klangliche Ergebnis sogar verschlechtert werden kann.

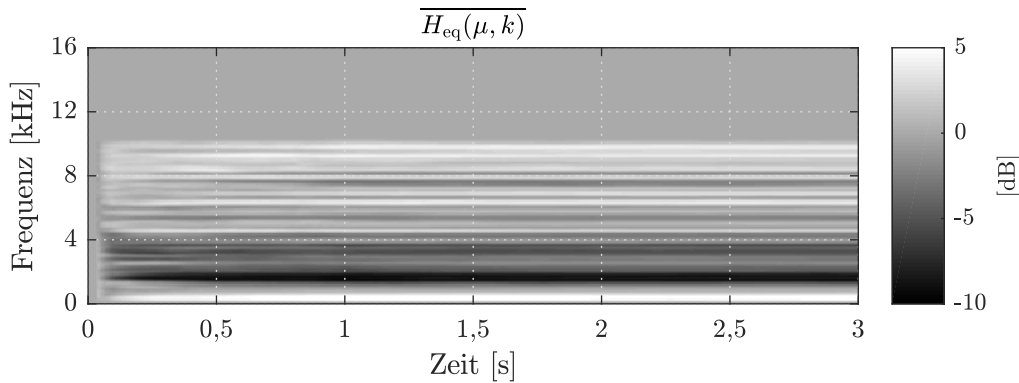


Abbildung 6.6: Der zeitliche Verlauf des Entzerrfilters als Spektrogramm.

In Abb. 6.6 sind die ersten drei Sekunden des Verlaufs der Filterkoeffizienten $\overline{H_{\text{eq}}(\mu, k)}$ als Spektrogramm gezeigt. Auch hier ist zu sehen, dass sich die Koeffizienten nur innerhalb der ersten 0,3s bis 0,5s ändern und danach konstant bleiben. Man erkennt außerdem, dass lediglich der Bereich bis 10 kHz entzerrt wird, darüber wird das Signal vom Entzerrer nicht verändert.

Abb. 6.7 zeigt die Betragsfrequenzgänge des Übertragungspfad zwischen Lautsprechern und Zuhöreremikrofon mit und ohne Entzerrung, welche sich nach 19,5s Simulationszeit einstellen. Der Frequenzgang nach der Entzerrung $\underline{H}_{\text{filt}}(k)$ wird berechnet durch Multiplikation des gemessenen Raumes mit dem Entzerrfilter

$$\underline{H}_{\text{filt}}(k) = \underline{H}_k(k) \cdot \overline{H_{\text{eq}}(k)}. \quad (6.10)$$

Insgesamt erkennt man, dass der entzerrte Frequenzgang (schwarz) bereits rein optisch im relevanten Frequenzbereich von 100 Hz bis 10 kHz deutlich linearer ist als vor der

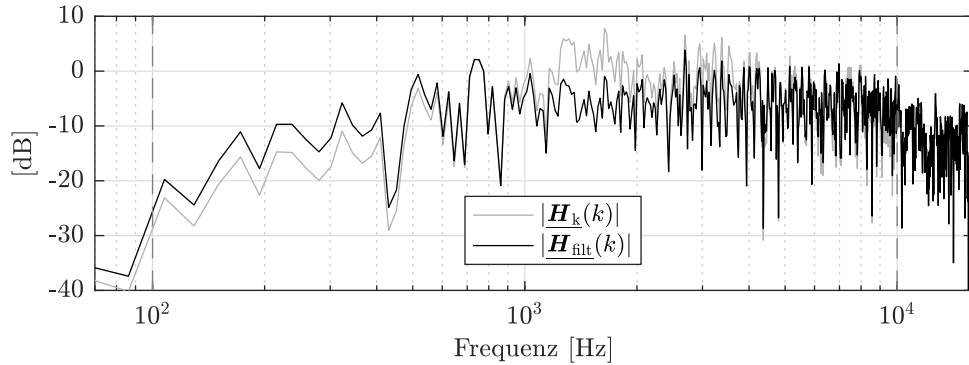


Abbildung 6.7: Der Frequenzgang am Zuhörerplatz mit und ohne Entzerrung nach 19,5 s Simulationszeit.

Entzerrung (grau). Es ist zu sehen, dass der tiefe Frequenzbereich von 100 Hz bis etwa 600 Hz um bis zu 5 dB angehoben wird, während umgekehrt die Überhöhungen im Bereich 1 Hz bis 2 kHz um etwa 8 dB bis 9 dB abgesenkt werden.

Diese Linearität kann durch ein quadratisches Abstandsmaß ausgedrückt werden. Hierfür wird die mittlere quadratische Distanz zum Mittelwert im Frequenzband μ_s berechnet. Für die Frequenzgänge $\underline{H}_k(k)$ und $\underline{H}_{\text{filt}}(k)$ erhält man auf diese Weise die Distanzen $D_k(k)$ und $D_{\text{filt}}(k)$, welche sich wie folgt berechnen

$$D_k(k) = \frac{1}{N_{\mu_s}} \sum_{\mu=\mu_u}^{\mu_o} \left(|\underline{H}_k(\mu, k)| - \frac{1}{N_{\mu_s}} \sum_{\mu=\mu_u}^{\mu_o} |\underline{H}_k(\mu, k)| \right)^2 \approx -7,45 \text{ dB}, \quad (6.11)$$

$$D_{\text{filt}}(k) = \frac{1}{N_{\mu_s}} \sum_{\mu=\mu_u}^{\mu_o} \left(|\underline{H}_{\text{filt}}(\mu, k)| - \frac{1}{N_{\mu_s}} \sum_{\mu=\mu_u}^{\mu_o} |\underline{H}_{\text{filt}}(\mu, k)| \right)^2 \approx -12,05 \text{ dB}. \quad (6.12)$$

Die Bandgrenzen des Intervalls μ_s werden wie oben auf den Frequenzbereich 100 Hz bis 10 kHz gesetzt. Berechnet man auf diese Weise die Distanzen nach 19,5 s Simulationszeit, erhält man für $D_k(k) = -7,45$ dB und für $D_{\text{filt}}(k) = -12,05$ dB für die verwendete Impulsantwort. In diesem Fall ist der quadratische Abstand zum Mittelwert mit automatischer Entzerrung also etwa 4,6 dB geringer als ohne Entzerrung.

6.1.4 Zusammenfassung

Es wurde ein Verfahren zur automatischen Entzerrung des Lautsprechersignals am Zuhörerohr vorgestellt. Das Verfahren wurde für den Einsatz in einer geschlossenen elektroakustischen Schleife mit Rückkopplungskompensator entwickelt. Das Ziel des Verfahrens ist es, am Ohr des Zuhörers einen flachen Frequenzgang einzustellen. Hierfür wird das Übertragungsverhalten zwischen Lautsprecher und Ohr mit einem adaptiven Filter geschätzt. Das Lautsprechersignal wird anschließend mit der Inversen des geschätzten Frequenzgangs gefiltert. Simulationen zeigen, dass Überhöhungen und Kerben im Frequenzgang durch das Verfahren abgeflacht werden. Subjektive Hörtests bestätigen dieses

Ergebnis. Die Sprache am Zuhörerohr klingt natürlicher und weniger spektral gefärbt als ohne die automatische Entzerrung.

6.2 Verbesserung der Sprachverständlichkeit

Nichtlinearitäten elektronischer Bauteile oder nichtlineare Signalverarbeitungsalgorithmen sind meist unerwünscht, da sie ein Signal verzerren. Im musikalischen Kontext gibt es jedoch Anwendungen, bei denen nichtlineare Systeme gezielt eingesetzt werden, um ein Signal mit seinen Harmonischen anzureichern. Als Harmonische werden ganzzahlige Vielfache eines Grundtons bezeichnet. Der Grundton ist dabei die erste Harmonische, die zweite Harmonische liegt bei der doppelten Frequenz des Grundtons, usw. Klanglich wird vor allem zwischen geradzahligen und ungeradzahligen Harmonischen unterschieden. So erzeugt beispielsweise ein Röhrenverstärker bei geringer Aussteuerung der Röhren vorrangig geradzahlige Harmonische. Diese sorgen dafür, dass der Klang eines Röhrenverstärkers subjektiv als *angenehm* und *warm* empfunden wird. Umgekehrt klingen die ungeradzahligen Harmonischen eher *hart* und *kalt*, was z. B. bei Verzerrern für die elektrische Gitarre ausgenutzt wird. Neben Röhrenverstärkern und Verzerrern sind es vor allem Effektgeräte wie der *Exciter* oder der artverwandte *Enhancer*, welche bei der Musikproduktion eingesetzt werden, um den Klang eines Audiosignals mittels absichtlicher Verzerrungen gezielt aufzuwerten. Prominentestes Beispiel ist der um 1970 von der Firma Aphex entwickelte *Aural Exciter* [1]. Aufgrund seiner Popularität wurden mögliche Anwendungsfälle für dieses Effektgerät untersucht, welche auch über die Musik hinaus gehen. So wird beispielsweise in [11] nachgewiesen, dass der *Aural Exciter* die Sprachverständlichkeit in geräuschbehafteter Umgebung signifikant verbessern kann. Da der *Exciter* im Original ein analoges Hardwaregerät ist, beschäftigt sich u. a. [60] mit der digitalen Implementierung des Effekts.

Auch im Rahmen dieser Arbeit wurde ein Algorithmus nach dem Vorbild des *Exciters* entwickelt, mit welchem die Sprachverständlichkeit bei ICC-Systemen verbessert werden kann. Dieser Algorithmus wurde in [10] veröffentlicht. In [25] wird die Idee dieses Ansatzes aufgegriffen und für Kommunikationseinheiten in Feuerschutzmasken adaptiert.

6.2.1 Prinzip des *Exciters*

Der prinzipielle Aufbau des *Exciters* ist in Abb. 6.8 gezeigt. In der geschlossenen Schleife eines ICC-Systems befindet sich der Algorithmus im Vorwärtspfad, d. h. als Eingangssignal dient das Fehlersignal des Rückkopplungskompensators $e(n)$, das Ausgangssignal ist das Lautsprechersignal $x(n)$. In Abb. 6.8 ist zu sehen, dass sich im Hauptpfad lediglich eine Verzögerung um D Abtastwerte befindet. Die Verzögerung gleicht die Laufzeit des Parallelpfads aus, damit der Hauptpfad phasenrichtig zum Parallelpfad addiert wird. Ohne diesen Ausgleich würde nach der Addition ein Kammfilter-Effekt entstehen. Im Parallelpfad wiederum werden die Harmonischen erzeugt. Der Vorteil der Parallelverarbeitung ist, dass der *Exciter*-Effekt gewichtet zum verzögerten Original addiert und damit die Stärke des Effekts eingestellt werden kann

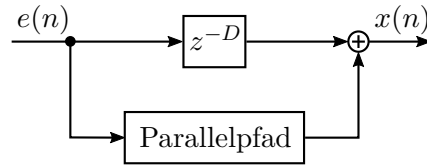


Abbildung 6.8: Prinzipielle Verarbeitung beim *Exciter*. Das Verzögerungsglied z^{-D} im Hauptpfad gleicht die Laufzeit des Parallelpfades aus.

Die detaillierten Verarbeitungsschritte sind in Abb. 6.9 dargestellt. Die grundlegende Funktionsweise des *Exciters* kann als Addition verschiedener Kennlinien aufgefasst werden. Zur Veranschaulichung dieses Prinzips ist in der Abbildung im Hauptpfad ei-

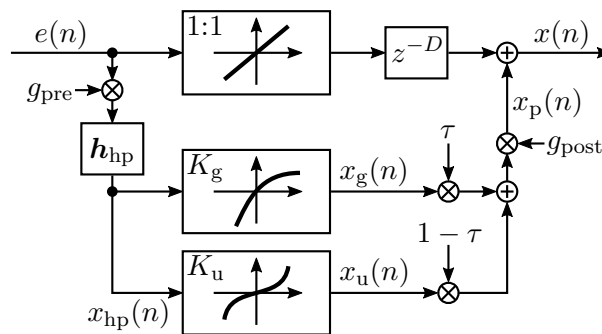


Abbildung 6.9: Blockschaltbild des *Exciters*. Im Parallelpfad werden mittels unterschiedlicher Kennlinien die Harmonischen erzeugt.

ne lineare Kennlinie eingezeichnet, welche das Signal unverarbeitet lässt, d. h. zwischen Ein- und Ausgang ist das Übertragungsverhältnis 1:1. Im Parallelpfad wird das Signal zunächst mit einem linearen Faktor g_{pre} gewichtet. Mit diesem kann eingestellt werden, wie stark die Kennlinien angesteuert werden, was unmittelbaren Einfluss auf die Stärke der Verzerrungen hat. Anschließend wird das Signal mit einem Hochpass h_{hp} gefiltert. Dieses Hochpassfilter ist notwendig, da tieffrequente Signalanteile mit hohen Pegeln die Kennlinien zu stark aussteuern würden. Die Harmonischen sollen vorwiegend aus den hohen Frequenzanteilen gebildet werden. Nach der Hochpassfilterung durchläuft das Signal die beiden nichtlinearen Kennlinien. Die Kennlinie K_g ist quadratisch und hat die Form

$$K_g : \quad x_g(n) = -\frac{1}{2}x_{\text{hp}}^2(n) + x_{\text{hp}}(n). \quad (6.13)$$

Diese Kennlinie erzeugt die zweite Harmonische einer Eingangsschwingung. Die ungeradzahigen Harmonischen werden durch eine Kennlinie K_u der Form

$$K_u \quad x_u(n) = |x_{\text{hp}}(n)| \cdot x_{\text{hp}}(n) \quad (6.14)$$

erzeugt. Aufgrund der Punktsymmetrie werden hiermit eine Vielzahl ungeradzahiger Harmonischer erzeugt. Das Mischungsverhältnis von geradzahigen und ungeradzahigen

Harmonischen wird mit einem Faktor τ eingestellt. Dieser Parameter bestimmt maßgeblich die Klangfarbe des *Exciters*. Da durch die Verzerrung auch Harmonische oberhalb der halben Abtastfrequenz erzeugt werden, treten Alias-Effekte auf. Diese können minimiert werden, indem das Signal vor der Verarbeitung mit der Kennlinie überabgetastet und anschließend wieder unterabgetastet wird. Der Übersicht halber ist dieser Schritt in Abb. 6.9 nicht eingezeichnet. Im Folgenden wird eine Über- bzw. Unterabtastung um den Faktor vier verwendet.

Bevor das Signal des Parallelpfads auf den zeitlich angeglichenen Hauptpfad addiert wird, wird es erneut mit einem linearen Faktor g_{post} gewichtet. Hiermit wird die Gesamtstärke des Effekts eingestellt. Für das Ausgangssignal des Parallelpfades $x_p(n)$ gilt folglich

$$x_p(n) = g_{\text{post}} \cdot (\tau \cdot x_g(n) + (1 - \tau) \cdot x_u(n)) \quad (6.15)$$

und damit für das Ausgangssignal, welches am Lautsprecher ausgegeben wird

$$x(n) = e(n - D) + x_p(n). \quad (6.16)$$

6.2.2 Parametrierung des Algorithmus

Im weiteren Verlauf werden zwei Parametersätze des *Exciters* untersucht. Die beiden mit A und B bezeichneten Varianten sind in Tab. 6.1 einander gegenübergestellt. Die Vari-

Parameter	Variante A	Variante B
Gewichtung Eingang	$g_{\text{pre}} = 1,8$	$g_{\text{pre}} = 4,0$
Gewichtung Ausgang	$g_{\text{post}} = 1,0$	$g_{\text{post}} = 1,0$
Grenzfrequenz Hochpass	$f_{\text{hp}} = 2 \text{ kHz}$	$f_{\text{hp}} = 2 \text{ kHz}$
Harmonische Gewichtung	$\tau = 0,8$	$\tau = 0,8$

Tabelle 6.1: Die Parameter der beiden getesteten Varianten.

anten unterscheiden sich in der Gewichtung g_{pre} , welche vor den Kennlinien angewendet wird. Der deutlich höhere Wert bei Variante B resultiert in stärkeren Verzerrungen. Die Grenzfrequenz des Hochpasses f_{hp} liegt in beiden Fällen bei 2 kHz.

Für den Hochpass wird ein Tiefen-Shelving-Filter mit 20 dB Bassdämpfung verwendet. Die Verzögerung im Vorwärtspfad entspricht der Gruppenlaufzeit dieses Filters. Der resultierende Frequenzgang des *Exciters* ist in Abb. 6.10 für die Variante A gezeigt. Zur Erstellung der Grafik wurden die beiden Nichtlinearitäten K_g und K_u durch die lineare 1:1 Kennlinie ersetzt. Der dargestellte Verlauf ergibt sich dann aus der Gewichtung des Parallelpfades mit $g_{\text{pre}} = 1,8$ und der Addition von Hauptpfad und Parallelpfad. Diese Höhenanhebung durch den Hochpass führt auch ohne nichtlineare Verzerrungen bereits zu einer Verbesserung der Sprachverständlichkeit. Der Grund hierfür ist, dass die für die Sprachverständlichkeit relevanten Informationen hauptsächlich bei hohen Frequenzen liegen. Diese Anteile werden vom vor allem tieffrequent energiereichen Fahrgeräusch maskiert [11]. Durch die Höhenanhebung werden diese Maskierungseffekte verringert. In

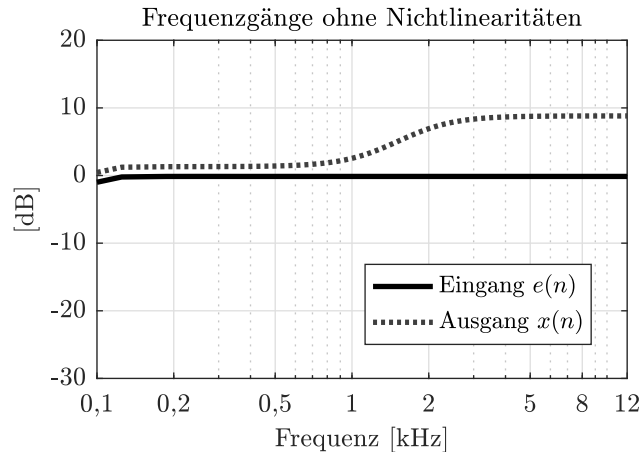


Abbildung 6.10: Der für weißes Eingangsrauschen am Ausgang gemessene Frequenzgang des *Exciters*. Für die Messung wurden die nichtlinearen Kennlinien durch die 1:1 Kennlinie ersetzt.

[11] wird jedoch nachgewiesen, dass die verbesserte Sprachverständlichkeit beim *Exciter* vorrangig auf die nichtlinearen Verzerrungen zurückzuführen ist.

Die beiden Kennlinien K_g nach Gl. (6.13) und K_u nach Gl. (6.14) sind in Abb. 6.11 dargestellt. Die schwarze Kennlinie K resultiert aus der Addition der Kennlinien

$$K = \tau \cdot K_g + (1 - \tau) \cdot K_u, \quad (6.17)$$

wobei für $\tau = 0,8$ gewählt wurde. Zum Vergleich ist die lineare 1:1 Kennlinie ebenfalls mit eingezeichnet.

Ein gängiges Maß zur Quantifizierung der harmonischen Verzerrung ist der Klirrfaktor (engl. *Harmonic Distortion*, HD). Für die i -te Harmonische berechnet sich der Klirrfaktor aus der Wurzel des Verhältnisses der Leistung P_i der i -ten Harmonischen zur Leistung P_1 der ersten Harmonischen, d. h.

$$\text{HD}_i = \sqrt{\frac{P_i}{P_1}}. \quad (6.18)$$

In Abb. 6.12 sind die Spektren der Ausgangssignale des *Exciters* für die Parameter der Varianten A und B gezeigt. Als Eingangssignal dient hier eine Sinusschwingung der Frequenz 1000 Hz und Amplitude 0,5. Der Hochpass im Parallelpfad ist nicht aktiv, d. h. $h_{\text{hp}} = 1$, dementsprechend ist die Verzögerung im Hauptpfad $D = 0$. Aus der Abbildung geht hervor, dass der Klirrfaktor für die zweite Harmonische bei der Variante A ca. 11,94 % beträgt. Bei den darüber liegenden ungeradzahigen Harmonischen nimmt der Klirrfaktor stetig ab. Ähnlich verhält es sich bei der Variante B, mit dem Unterschied, dass durch die größere Verstärkung am Eingang die Kennlinien deutlich stärker angesteuert werden. Hierdurch weist die zweite Harmonische einen Klirrfaktor von ca. 28,79 % auf und auch die ungeradzahigen Harmonischen liegen etwa im gleichen Verhältnis über den Klirrfaktoren der Variante A.

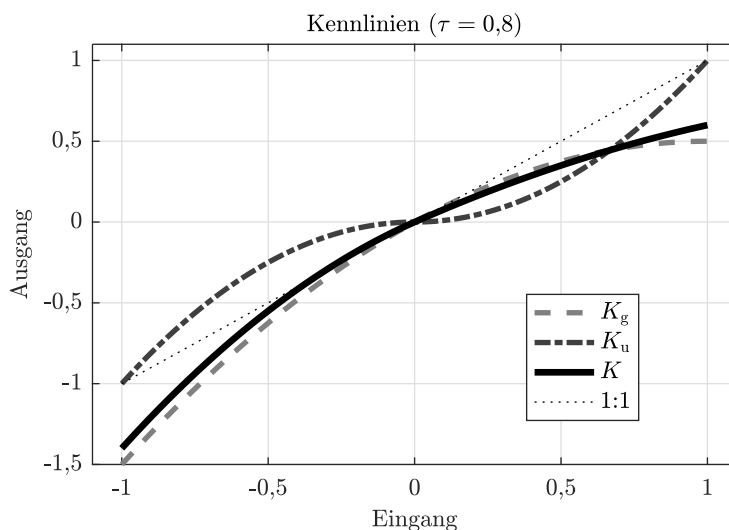


Abbildung 6.11: Die Kennlinien K_g und K_u sowie die aus der Addition resultierende Kennlinie K bei einem Mischungsverhältnis von $\tau = 0,8$.

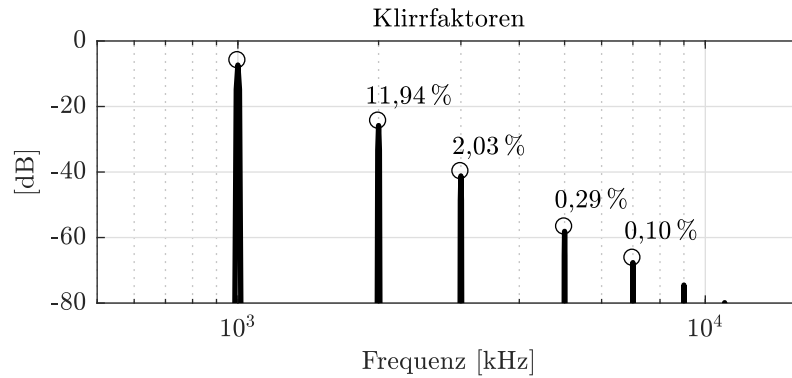
6.2.3 Ergebnis und Evaluierung

Der beschriebene Exciteralgorithmus wird in [10] hinsichtlich der erreichten Sprachqualität und Sprachverständlichkeit mit einem *Comparison Mean Opinion Score* (CMOS) Test evaluiert [33]. Beim CMOS Test handelt es sich um ein auditives Verfahren zur Beurteilung von Sprachqualität oder Sprachverständlichkeit, d. h. menschliche Hörer beurteilen die Signale nach ihrem persönlichen Empfinden. Neben den auditiven Tests gibt es verschiedene instrumentelle Verfahren, welche allerdings in der vorliegenden Arbeit nicht zum Einsatz kamen. Ein umfassender Überblick über Sprachqualitätsmessungen wird in [18] gegeben.

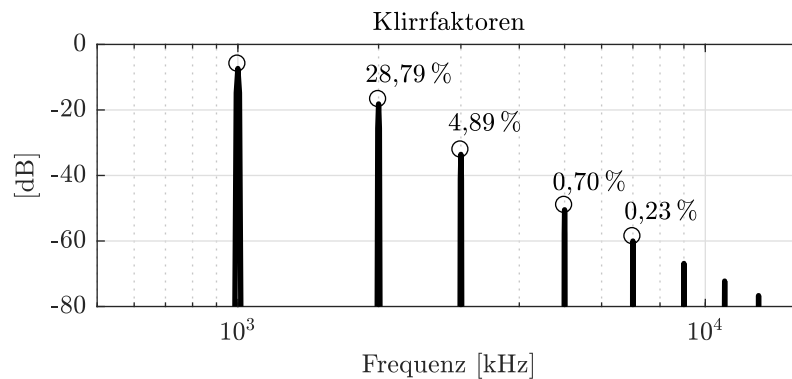
Der Test wird auf zwei grundsätzliche Fragestellungen ausgerichtet. Die erste Fragestellung ist, ob die Sprachqualität für Sprache mit bereits guter Qualität noch weiter verbessert werden kann. Dabei werden die Parameter des Verfahrens so gewählt, dass das Sprachsignal zwar hörbar aber noch nicht zu stark verändert wird. In diesem Bereich ist die Sprachverständlichkeit nicht beeinträchtigt, sie beträgt also stets 100 %. Die zweite grundsätzliche Frage ist, ob mit dem *Exciter* die Sprachverständlichkeit erhöht werden kann. Hier muss die Sprachverständlichkeit des unverarbeiteten Sprachsignals zunächst deutlich unter 100 % liegen, um eine Verbesserung feststellen zu können. Die Parameter für den *Exciter* werden in ihrer Wirkung aggressiver gewählt. Die Sprachqualität ist hier von untergeordneter Bedeutung.

Die Sprachqualität wird mit dem Parametersatz der Variante A in drei verschiedenen Szenarien evaluiert:

- S1: Ein klares Sprachsignal ohne Hintergrundgeräusch wird mit dem *Exciter* verarbeitet.



(a) Variante A.



(b) Variante B.

Abbildung 6.12: Die Klirrfaktoren der Harmonischen, gemessen mit einem Sinussignal der Frequenz 1 kHz und Amplitude 0,5.

- S2: Ein Sprachsignal wird von Fahrgeräusch bei 140 km/h überlagert. Das geräuschbehaftete Signal erfährt zunächst eine ideale Geräuschreduktion um -14 dB, bevor es mit dem *Exciter* verarbeitet wird. In einem ICC-System entspricht dieses Signal dem Lautsprechersignal bzw. bei Freisprecheinrichtungen dem Uplink.
- S3: Ein klares Sprachsignal wird zunächst mit dem *Exciter* verarbeitet und anschließend von Fahrgeräusch bei 140 km/h überlagert. Im ICC-System entspricht dies einem Lautsprechersignal, welches in geräuschbehafteter Umgebung wiedergegeben wird bzw. bei der Freisprecheinrichtung dem Downlink.

Die Sprachverständlichkeit wird mit dem Parametersatz der Variante B in dem nachfolgend beschriebenen vierten Szenario evaluiert:

- S4: Dieses Szenario ist ähnlich zu S3 mit dem Unterschied, dass Fahrgeräusch bei 190 km/h verwendet wird. Zudem wird der Sprachpegel vor der Überlagerung mit dem Fahrgeräusch soweit abgesenkt, dass die Sprachverständlichkeit zwischen ca. 70 % und 90 % liegt.

Im Hörtest wird jedes Szenario jeweils mit aktivem *Exciter* gegen das gleiche Szenario ohne *Exciter* verglichen. Aus diesem Grund lautet die Fragestellung bei der Evaluierung der Sprachqualität „*Bewerten Sie, welches der beiden Sprachsignale besser klingt*“ und dementsprechend für die Sprachverständlichkeit „*Bewerten Sie, bei welchem der beiden Sprachsignale die Sprachverständlichkeit besser ist*“. Um einen Vergleich zu ermöglichen, werden etwaige Pegelunterschiede zwischen verarbeitetem und unverarbeitetem Signal ausgeglichen, sodass die Signale A-gewichtet den gleichen Pegel haben. Die beiden Signale können schließlich mit der folgenden dreistufigen CMOS Skala bewertet werden:

- Signal a) wird bevorzugt.
- Kein Unterschied hörbar.
- Signal b) wird bevorzugt.

Die Probanden wissen nicht, welches der beiden Signale a) und b) mit dem *Exciter* verarbeitet wurde und welches nicht. Jedes der vier Szenarien wird jeweils mit fünf verschiedenen Sprechern präsentiert. Zwei der Sprecher sind männlich, drei weiblich. Jeder Sprecher sagt zwei kurze, aufeinander folgende Sätze. Die Signale werden in zufälliger Reihenfolge präsentiert.

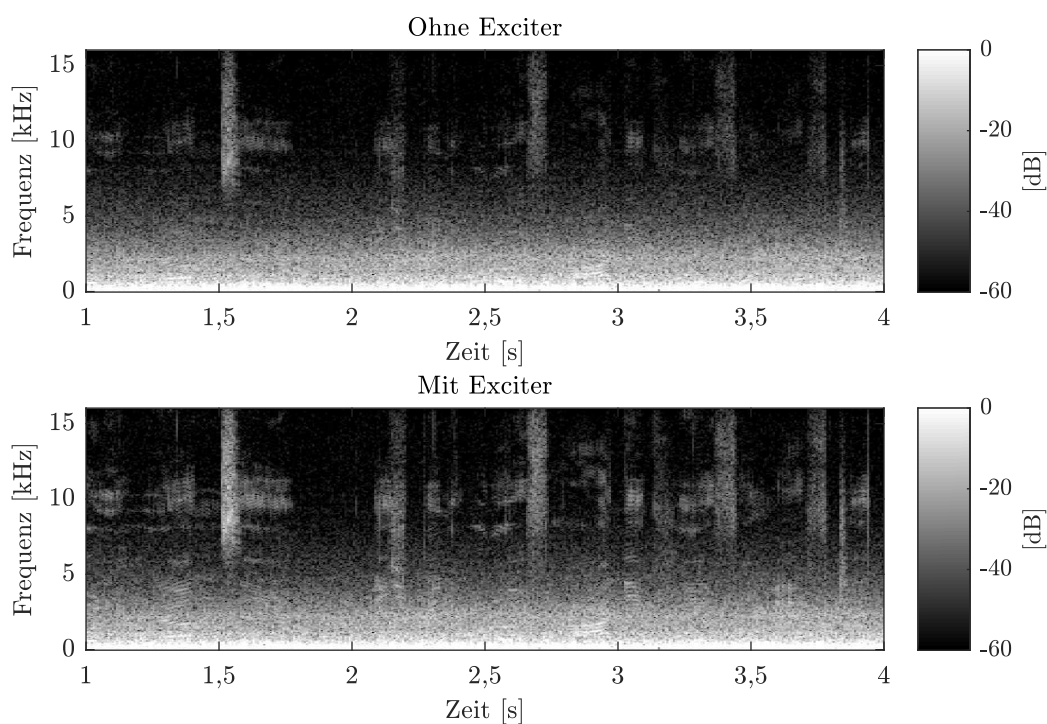


Abbildung 6.13: Spektrogramme einer weiblichen Sprecherin aus dem Szenario S4 mit und ohne *Exciter*.

In Abb. 6.13 ist beispielhaft ein Ausschnitt aus dem Spektrogramm einer weiblichen Sprecherin aus Szenario S4 gezeigt. Es ist zu erkennen, dass mit und ohne *Exciter* tief-

frequent das Hintergrundgeräusch jeweils stark dominiert. Hingegen ist bei den hohen Frequenzen ab ca. 2 kHz zu sehen, dass nach der Verarbeitung mit dem *Exciter* in diesem Frequenzbereich deutlich mehr Energie vorhanden ist, als ohne *Exciter*.

Die Audiodaten werden den Probanden über Kopfhörer präsentiert. Insgesamt nahmen an dem Hörtest 18 normalhörende Probanden im Alter von 22 bis 44 Jahren teil, davon 4 weiblich und 14 männlich.

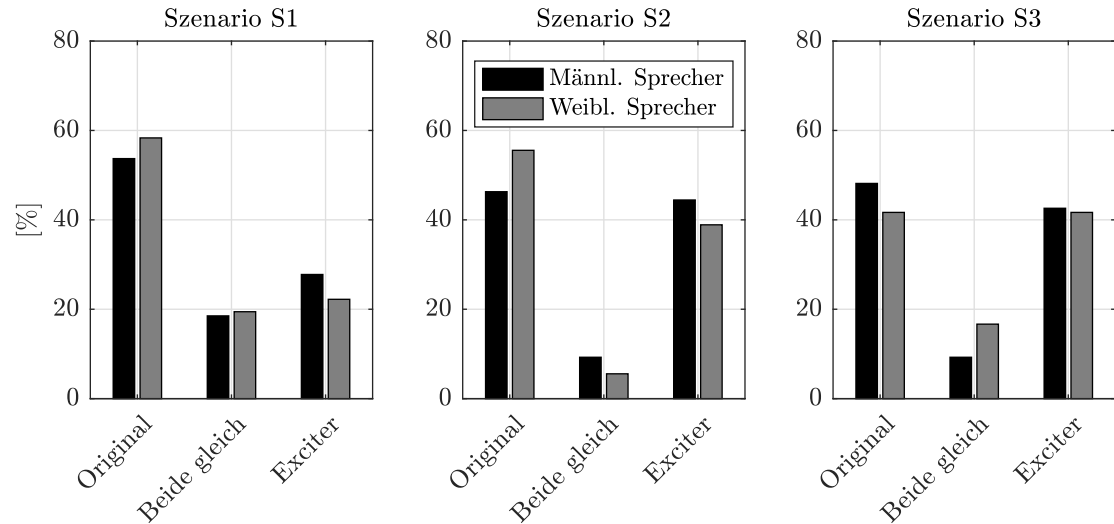


Abbildung 6.14: Das Ergebnis der drei Szenarien zur Evaluierung der Sprachqualität.

Das Ergebnis der Szenarien zur Evaluierung der Sprachqualität in den Szenarien S1, S2 und S3 ist in Abb. 6.14 gezeigt. Da der *Exciter* aufgrund des Hochpasses vor allem die hochfrequenten Anteile beeinflusst, werden die Ergebnisse für die männlichen und die weiblichen Sprecher getrennt dargestellt.

In Szenario S1 erkennt man, dass die Verarbeitung der klaren Sprachsignale mit dem *Exciter* im vorliegenden Fall keine Verbesserung der Sprachqualität bringt. Von den Probanden bevorzugen sowohl bei den männlichen Sprechern als auch bei den weiblichen Sprechern jeweils über 50 % das unverarbeitete Signal (in der Abbildung als *Original* bezeichnet). Lediglich rund 28 % bevorzugen den *Exciter* bei den männlichen Sprechern und 22 % bei den weiblichen Sprechern. Knapp 20 % nehmen keinen Unterschied zwischen den beiden Varianten wahr.

In Szenario S2 zeigt sich ebenfalls, dass die Mehrheit der Probanden das unverarbeitete Signal gegenüber dem *Exciter* bevorzugt. Allerdings wird hier der *Exciter* besser bewertet als in Szenario S1. So bevorzugen den *Exciter* ca. 44 % bei den männlichen Sprechern und ca. 39 % bei den weiblichen.

Ähnlich ausgewogen ist das Bild im dritten Szenario S3. Hier bevorzugen bei den weiblichen Sprechern jeweils 42 % der Probanden den *Exciter* und 42 % das unverarbeitete Signal. Die verbleibenden etwa 16 % können keinen Unterschied wahrnehmen. Bei den männlichen Sprechern liegt abermals das unverarbeitete Signal mit 48 % vor dem *Exciter* (ca. 43 %).

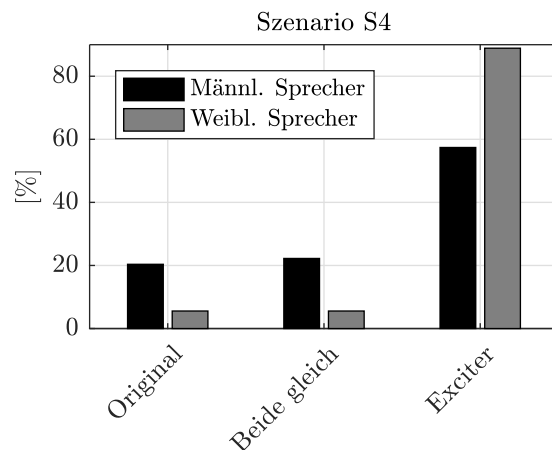


Abbildung 6.15: Das Ergebnis der Evaluierung des Szenarios S4 zur Ermittlung der Sprachverständlichkeit bei der Wiedergabe in einer sehr lauten Geräuschumgebung.

Deutlich besser fällt das Ergebnis zugunsten des *Exciters* aus, wenn man die Evaluierung der Sprachverständlichkeit betrachtet. Dieses Ergebnis ist in Abb. 6.15 gezeigt. Hier geben etwa 89% der Probanden an, die weiblichen Sprecher mit *Exciter* besser zu verstehen, als ohne (ca. 6%). Auch bei den männlichen Sprechern gibt die Mehrheit von ca. 57% an, dass die Sprecher mit *Exciter* besser verständlich sind als ohne (ca. 20%).

6.2.4 Zusammenfassung

Es wurde ein Verfahren vorgestellt, mit welchem ein Signal mit seinen Harmonischen angereichert werden kann. Hierfür durchläuft das Signal zunächst zwei unterschiedliche, nichtlineare Kennlinien. Anschließend wird das mit den Kennlinien verarbeitete Signal gewichtet zum unverarbeiteten addiert. Hörtest zeigten, dass durch das Hinzufügen von Harmonischen die Sprachverständlichkeit eines Sprachsignals bei der Wiedergabe in einer stark geräuschbehafteten Umgebung deutlich verbessert werden kann. Subjektive Tests im Demonstratorfahrzeug bestätigen dieses Ergebnis. Eine Verbesserung der Sprachqualität konnte in dem Hörtest nicht nachgewiesen werden. Letzteres kann an der bereits sehr hohen Qualität der Testsignale liegen.

Kapitel 7

Zusammenfassung und Ausblick

Im Rahmen der vorliegenden Arbeit wurden verschiedene Algorithmen zur Verbesserung der Stabilität von ICC-Systemen entwickelt. Darüber hinaus wurden Verfahren beschrieben, mit welchen die Signale klanglich aufgewertet können und zudem auch die Sprachverständlichkeit erhöht werden kann. Die Algorithmen wurden zunächst theoretisch untersucht, anschließend wurde die Funktionsweise in Simulationen verifiziert und optimiert. Um die Algorithmen im realen Betrieb testen zu können, wurde ein Demonstratorfahrzeug aufgebaut. Die akustischen Eigenschaften dieses Fahrzeugs wurden in Messungen ermittelt und auf Basis der Messungen die Anforderungen an die Signalverarbeitung abgeleitet. Für die Echtzeittests im Demonstratorfahrzeug wurden die Algorithmen in das eigens entwickelte Rahmenwerk RACE implementiert. Dieses Rahmenwerk wurde in zahlreichen Testfahrten erfolgreich eingesetzt, um die Performanz der ICC-Systeme zu präsentieren. Zahlreiche weitere Algorithmen, die in dieser Arbeit zwar nicht beschrieben werden aber für den Betrieb von ICC-Systemen notwendig sind, wurden im Rahmen studentischer Arbeiten entwickelt und in RACE integriert. Hierzu zählen beispielsweise Geräuschreduktionen, Equalizer oder Kompressoren (vgl. Anhang B).

7.1 Zusammenfassung

Mit der in dieser Arbeit vorgestellten Schrittweitensteuerung für die adaptive Rückkopplungskompensation wird es ermöglicht, dass der NLMS-Algorithmus auch in einer geschlossenen elektro-akustischen Schleife mit stark korrelierten Signalen konvergieren kann. Dadurch können hohe Verstärkungen erzielt werden, ohne dass das System instabil wird. Dies resultiert zugleich auch in einer besseren Klangqualität, da Rückkopplungsartefakte kompensiert werden und damit nicht hörbar sind. Da die Schrittweitensteuerung kurze Zeitpunkte ausnutzt, in denen die Signale nicht korreliert sind, sind im Gegensatz zu gängigen Kompensationsansätzen keine weitere Dekorrelationsmaßnahmen notwendig. Neben einer geringeren Rechenlast hat dies zudem den Vorteil einer natürlicheren Klangqualität. Der Grund ist, dass die Dekorrelation in den Signalpfad integriert werden muss, wodurch der Klang des Lautsprechersignals zwangsläufig beeinflusst wird. Die Herausforderung bei der vorgestellten Schrittweitensteuerung ist die Schätzung des Systemabstands und die damit zusammenhängende Detektion von Raumänderungen. Hierfür wurden verschiedene Verfahren entwickelt und untersucht. Diese sind in der Lage, auch bei starken Raumänderungen den Systemabstand korrekt zu schätzen und damit Stabilität zu gewährleisten.

Je geringer die verfügbare Rechenleistung, desto kürzer ist der Teil der Impulsantwort, der mit dem Rückkopplungskompensator abgebildet werden kann. Dies kann dazu führen, dass der Kompensator nicht den kompletten relevanten Teil der Impulsantwort abdeckt. Durch die nicht kompensierten späten Anteile der Impulsantwort können Restrückkopplungen hörbar werden, die zum einen das Klangergebnis beeinträchtigen, zum anderen aber auch das System zum Aufschwingen bringen können. Um diese Restrückkopplungen zu unterdrücken, wurde ein Postfilter entwickelt, welches abhängig von der Länge und vom Abgleichzustand des Kompensators Restrückkopplungen mit einem Wiener-Filter unterdrückt. Hierdurch kann der Klangeindruck deutlich verbessert werden.

Eine weitere klangliche Verbesserung wurde durch eine automatische Entzerrung des Lautsprecher Signals am Zuhörerplatz erreicht. Hierbei werden Überhöhungen, bzw. Vertiefungen im Frequenzgang, die beispielsweise vom Raum oder den Lautsprechern verursacht werden, abgeflacht. Nach der automatischen Entzerrung entsteht beim Zuhörer der Eindruck eines flachen Frequenzgangs.

Die Sprachverständlichkeit des über die Lautsprecher des ICC-Systems wiedergegebenen Sprachsignals ist aufgrund der Hintergrundgeräusche im Fahrzeug oftmals beeinträchtigt. In dieser Arbeit konnte nachgewiesen werden, dass die Sprachverständlichkeit durch gezieltes Hinzufügen von Harmonischen in bestimmten Szenarien verbessert werden kann. Das vorgestellte Verfahren basiert auf dem oftmals in der Musikproduktion verwendeten *Exciter*.

Sämtliche Algorithmen wurden im Hinblick auf eine echtzeitfähige Implementierung entwickelt. Hierbei wurde darauf geachtet, dass die benötigte Rechenleistung gering gehalten wird, was beispielsweise durch die Verarbeitung im Teilband mittels Filterbänken realisiert wurde. Damit für die verschiedenen Algorithmen jeweils die optimale Filterbank zum Einsatz kommt, wurde eine Struktur aus zwei verschiedenen Filterbänken implementiert. Gleichzeitig wurde die Verzögerung im Signalpfad durch entsprechende Blockgrößen der Filterbänke unter 15 ms gehalten, um das klangliche Ergebnis nicht zu beeinträchtigen.

Während vieler Testfahrten mit den Demonstratorfahrzeugen konnte das in dieser Arbeit entwickelte ICC-System im realen Betrieb erlebt werden. Hierbei konnte festgestellt werden, dass die beiden eingangs gesetzten Ziele, nämlich die Vergrößerung der maximalen stabilen Verstärkung bei gleichzeitig verbesserter Klangqualität, mit den vorgestellten Algorithmen erreicht werden. Auch bei Verstärkungen, die deutlich über denen auf dem Markt verfügbarer ICC-Systeme liegen, arbeitet das System zuverlässig und stabil. Dabei bleibt der Höreindruck stets natürlich ohne störende Artefakte.

7.2 Ausblick

Für den Einsatz des ICC-Systems in Serienfahrzeugen wird die benötigte Rechenleistung weiter optimiert werden müssen, da die Algorithmen hierfür auf einem eingebetteten System mit begrenzter Rechenkapazität implementiert werden. Dies kann u. a. dadurch erreicht werden, dass sowohl für die Module als auch den Rückkopplungskompensator ei-

ne einheitliche Filterbank verwendet wird. Da aufgrund fehlender Projektionen *Overlap-Add*-Filterbänke mit einem gleitenden Fenster effizienter zu berechnen sind als *Overlap-Save*-Filterbänke, kommen in Serienprodukten hauptsächlich erstere zum Einsatz. Hier gilt es zu untersuchen, wie sich die schrittweitengesteuerte Rückkopplungskompensation innerhalb einer solchen Filterbank verhält. Aufgrund der fehlenden Projektionen kann es sein, dass die Schrittweitensteuerung den Nachhall hier weniger präzise detektiert. Da gewichtete *Overlap-Add*-Filterbänke die zyklischen Faltungsfehler nicht vollständig unterdrücken, wird dies zudem Auswirkungen auf das Konvergenzverhalten des adaptiven Filters haben.

Die hier vorgestellten Algorithmen wurden sowohl für Einwege- als auch bidirektionale ICC-Systeme verwendet. Im Fall der bidirektionalen Systeme kommt eine Pegelwaage zum Einsatz, welche den inaktiven Weg stumm schaltet, sodass immer nur ein Weg zur gleichen Zeit aktiv ist. Schaltet die Pegelwaage schnell und präzise zwischen den beiden Wegen um, stellt dies für eine natürliche Kommunikation kein Problem dar. Ein gleichzeitiger Betrieb von zwei Wegen ist nur eingeschränkt möglich, da das System hier bereits bei geringen Verstärkungen instabil wird. Der Grund ist, dass das System zusätzlich zu den langen Pfaden auch über die kurzen Pfade zurück gekoppelt ist. Soll es dennoch ermöglicht werden, dass beide Wege gleichzeitig aktiv sind, muss in zukünftigen Arbeiten die Schrittweitensteuerung des adaptiven Filters weiterentwickelt werden.

Das Verfahren für die automatische Entzerrung bietet ebenfalls Raum für weitere Verbesserungen, die im Rahmen dieser Arbeit nicht untersucht wurden. Beispielsweise könnte man das Entzerrfilter mit der spektralen Einhüllenden des Hintergrundgeräusches gewichten. Dadurch wird die Sprache in stark geräuschbehafteten Bändern stärker angehoben, was sich in einem verbesserten SNR bemerkbar macht. Alternativ könnte man als Wunschfrequenzgang auch den Frequenzgang vom Mund des Sprechers zum Zuhörerohr ohne Rückkopplung vorgeben. Dafür müsste man diesen Frequenzgang einmalig vorab messen. Auf diese Weise ist die Sprache am Ohr spektral genauso gefärbt wie der Direktschall ohne ICC-System. Es ist zu untersuchen, ob dies in einem noch natürlicheren Klangbild resultiert. Zur Verbesserung der Stabilität könnte man zudem die Überhöhungen des Rückkopplungspfades mit dem Entzerrer dämpfen. Dabei muss untersucht werden, ob die Resonanzen dieses Pfades nahe bei denen des kurzen Pfades liegen, da es sich um den gleichen Raum handelt. Ein solcher Mechanismus wäre nur notwendig, wenn sich die Resonanzen von langem und kurzem Pfad unterscheiden würden.

Mit dem *Exciter* konnte zwar die Sprachverständlichkeit verbessert werden, eine Verbesserung der subjektiv empfundenen Sprachqualität konnte jedoch nicht nachgewiesen werden. Normalerweise wird der *Exciter* in Musikproduktionen angewendet, wo es seine Aufgabe ist, die Signale klanglich aufzuwerten. Hier könnte man in weiterführenden Arbeiten untersuchen, weshalb dies bei den im Hörtest verwendeten Sprachsignalen nicht der Fall war. Ein Grund könnte sein, dass die im Hörtest verwendeten Sprachdaten bereits von Grund auf eine hohe Qualität aufweisen, sodass jede zusätzliche Verarbeitung die Natürlichkeit beeinträchtigt. Dies entspricht der Tatsache, dass die Qualität des *Exciters* besser bewertet wird, wenn das Sprachsignal vom Hintergrundgeräusch gestört wird, während beim klaren Sprachsignal ohne Hintergrundgeräusch der *Exciter* am schlechtesten bewertet wird. In weiterführenden Arbeiten muss daher untersucht werden,

ob es mit dem *Exciter* möglich ist, Signale von minderer Qualität, wie sie im Fahrzeug vorkommen können, aufzuwerten. In diesem Zusammenhang könnte der *Exciter* auch als eine Art Bandbreitenerweiterung für bandbegrenzte Signale untersucht werden.

Anhang A

Herleitung der optimalen Schrittweite

In Kap. 4 wird die Konvergenz des adaptiven Filters für die Rückkopplungskompensation mittels Schrittweitensteuerung verbessert. Die detaillierte Herleitung der optimalen Schrittweite wird aus Gründen der besseren Lesbarkeit an dieser Stelle beschrieben. Die Herleitung ist angelehnt an [29, 31]. Dort wird die optimale Schrittweite für Signale im Zeitbereich hergeleitet. Die nachfolgenden Überlegungen werden zwar im Teilband durchgeführt, basieren jedoch auf den genannten Quellen.

Ausgangslage ist, wie in Abschn. 4.1 erläutert, Gl. (4.1)

$$\hat{\mathbf{H}}(\mu, k+1) = \hat{\mathbf{H}}(\mu, k) + \alpha(\mu, k) \cdot \frac{E(\mu, k) \cdot \mathbf{X}^*(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}. \quad (\text{A.1})$$

Diese Gleichung beschreibt die rekursive Aktualisierung der Filterkoeffizienten $\hat{\mathbf{H}}(\mu, k)$. Die Differenz zwischen den Koeffizienten von realer Teilbandimpulsantwort $\mathbf{H}(\mu)$ und geschätzter Teilbandimpulsantwort $\hat{\mathbf{H}}(\mu, k)$ ist

$$\mathbf{H}_{\Delta}(\mu, k) = \mathbf{H}(\mu) - \hat{\mathbf{H}}(\mu, k). \quad (\text{A.2})$$

Setzt man Gl. (A.1) in Gl. (A.2) ein, so erhält man die rekursive Berechnung der Differenz

$$\mathbf{H}_{\Delta}(\mu, k+1) = \mathbf{H}_{\Delta}(\mu, k) - \alpha(\mu, k) \cdot \frac{E(\mu, k) \cdot \mathbf{X}^*(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}. \quad (\text{A.3})$$

Der Systemabstand der Teilbandimpulsantwort wird berechnet aus der quadratischen Norm von (A.3)

$$\|\mathbf{H}_{\Delta}(\mu, k+1)\|^2 = \sum_{m=0}^{M-1} |H_{\Delta m}(\mu, k+1)|^2, \quad (\text{A.4})$$

$$|H_{\Delta m}(\mu, k+1)|^2 = H_{\Delta m}^*(\mu, k+1) \cdot H_{\Delta m}(\mu, k+1). \quad (\text{A.5})$$

Für Gl. (A.3) folgt damit

$$\|\mathbf{H}_{\Delta}(\mu, k+1)\|^2 = \|\mathbf{H}_{\Delta}(\mu, k) - \alpha(\mu, k) \cdot \frac{E(\mu, k) \cdot \mathbf{X}^*(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}\|^2 \quad (\text{A.6})$$

$$= \sum_{m=1}^{M-1} \left(H_{\Delta m}^*(\mu, k) - \alpha(\mu, k) \cdot \frac{E^*(\mu, k) \cdot X(\mu, k-m)}{\|\mathbf{X}(\mu, k)\|^2} \right) \cdot \left(H_{\Delta m}(\mu, k) - \alpha(\mu, k) \cdot \frac{E(\mu, k) \cdot X^*(\mu, k-m)}{\|\mathbf{X}(\mu, k)\|^2} \right). \quad (\text{A.7})$$

Nach Ausmultiplizieren erhält man

$$\begin{aligned}
 \|\mathbf{H}_\Delta(\mu, k+1)\|^2 &= \sum_{m=0}^{M-1} H_{\Delta m}^*(\mu, k) \cdot H_{\Delta m}(\mu, k) \\
 &+ \alpha^2(\mu, k) \cdot \frac{E^*(\mu, k) \cdot X(\mu, k-m) \cdot E(\mu, k) \cdot X^*(\mu, k-m)}{\|\mathbf{X}(\mu, k)\|^4} \\
 &- H_{\Delta m}^*(\mu, k) \cdot \alpha(\mu, k) \cdot \frac{E(\mu, k) \cdot X^*(\mu, k-m)}{\|\mathbf{X}(\mu, k)\|^2} \\
 &- \alpha(\mu, k) \cdot \frac{E^*(\mu, k) \cdot X(\mu, k-m)}{\|\mathbf{X}(\mu, k)\|^2} \cdot H_{\Delta m}(\mu, k). \tag{A.8}
 \end{aligned}$$

Ersetzt man die Summation durch Vektorschreibweise, vereinfacht sich die Gleichung zu

$$\begin{aligned}
 \|\mathbf{H}_\Delta(\mu, k+1)\|^2 &= \|\mathbf{H}_\Delta(\mu, k)\|^2 + \alpha^2(\mu, k) \cdot \frac{|E(\mu, k)|^2 \cdot \mathbf{X}^T(\mu, k) \cdot \mathbf{X}^*(\mu, k)}{\|\mathbf{X}(\mu, k)\|^4} \\
 &- \alpha(\mu, k) \cdot \frac{E(\mu, k) \mathbf{H}_\Delta^H(\mu, k) \mathbf{X}^*(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2} \\
 &- \alpha(\mu, k) \cdot \frac{E^*(\mu, k) \mathbf{X}^T(\mu, k) \mathbf{H}_\Delta(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}. \tag{A.9}
 \end{aligned}$$

Setzt man den ungestörten Fehler nach Gl. (4.9)

$$E_u(\mu, k) = \mathbf{X}^T(\mu, k) \cdot \mathbf{H}_\Delta(\mu, k) \tag{A.10}$$

ein, erhält man nach Vereinfachen

$$\begin{aligned}
 \|\mathbf{H}_\Delta(\mu, k+1)\|^2 &= \|\mathbf{H}_\Delta(\mu, k)\|^2 + \alpha^2(\mu, k) \cdot \frac{|E(\mu, k)|^2}{\|\mathbf{X}(\mu, k)\|^2} \\
 &- \alpha(\mu, k) \cdot \frac{E(\mu, k) E_u^*(\mu, k) + E^*(\mu, k) E_u(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}. \tag{A.11}
 \end{aligned}$$

Ziel ist es, dass der Systemabstand nach Gl. (A.5) mit jedem Zeitschritt k kleiner wird, d. h.

$$\mathbb{E}\left\{\|\mathbf{H}_\Delta(\mu, k+1)\|^2\right\} - \mathbb{E}\left\{\|\mathbf{H}_\Delta(\mu, k)\|^2\right\} \leq 0. \tag{A.12}$$

Aus Gl. (A.11) folgt damit

$$\alpha^2(\mu, k) \cdot \mathbb{E}\left\{\frac{|E(\mu, k)|^2}{\|\mathbf{X}(\mu, k)\|^2}\right\} - \alpha(\mu, k) \cdot \mathbb{E}\left\{\frac{E(\mu, k) E_u^*(\mu, k) + E^*(\mu, k) E_u(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}\right\} \leq 0, \tag{A.13}$$

was bedeutet, dass die Schrittweite im Intervall

$$0 \leq \alpha(\mu, k) \leq \frac{\mathbb{E}\left\{\frac{E(\mu, k) E_u^*(\mu, k) + E^*(\mu, k) E_u(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}\right\}}{\mathbb{E}\left\{\frac{|E(\mu, k)|^2}{\|\mathbf{X}(\mu, k)\|^2}\right\}} \tag{A.14}$$

liegt. Leitet man den Systemabstand (A.11) nach der Schrittweite ab und setzt die Ableitung gleich null, erhält man die optimale Schrittweite

$$\frac{\partial \mathbb{E}\{\|\mathbf{H}_\Delta(\mu, k+1)\|^2\}}{\partial \alpha(\mu, k)} \stackrel{!}{=} 0. \quad (\text{A.15})$$

Auflösen von

$$0 = 2 \cdot \alpha_{\text{opt}}(\mu, k) \cdot \mathbb{E}\left\{\frac{|E(\mu, k)|^2}{\|\mathbf{X}(\mu, k)\|^2}\right\} - \mathbb{E}\left\{\frac{E(\mu, k)E_u^*(\mu, k) + E^*(\mu, k)E_u(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}\right\} \quad (\text{A.16})$$

nach der Schrittweite $\alpha_{\text{opt}}(\mu, k)$ führt schließlich auf

$$\alpha_{\text{opt}}(\mu, k) = \frac{1}{2} \cdot \frac{\mathbb{E}\left\{\frac{E(\mu, k)E_u^*(\mu, k) + E^*(\mu, k)E_u(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}\right\}}{\mathbb{E}\left\{\frac{|E(\mu, k)|^2}{\|\mathbf{X}(\mu, k)\|^2}\right\}}. \quad (\text{A.17})$$

Der gestörte Fehler $E(\mu, k)$ kann berechnet werden aus der Überlagerung von ungestörtem Fehler und lokalem Störer

$$E(\mu, k) = E_u(\mu, k) + S(\mu, k). \quad (\text{A.18})$$

Eingesetzt in Gl. (A.17) ergibt sich

$$\begin{aligned} \alpha_{\text{opt}}(\mu, k) &= \frac{1}{2} \cdot \frac{\mathbb{E}\left\{\frac{(E_u(\mu, k) + S(\mu, k))E_u^*(\mu, k) + (E_u^*(\mu, k) + S^*(\mu, k))E_u(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}\right\}}{\mathbb{E}\left\{\frac{|E(\mu, k)|^2}{\|\mathbf{X}(\mu, k)\|^2}\right\}} \\ &= \frac{1}{2} \cdot \frac{\mathbb{E}\left\{\frac{2|E_u(\mu, k)|^2 + S(\mu, k)E_u^*(\mu, k) + S^*(\mu, k)E_u(\mu, k)}{\|\mathbf{X}(\mu, k)\|^2}\right\}}{\mathbb{E}\left\{\frac{|E(\mu, k)|^2}{\|\mathbf{X}(\mu, k)\|^2}\right\}}. \end{aligned} \quad (\text{A.19})$$

Die lokale Störung $S(\mu, k)$ und der ungestörte Fehler $E_u(\mu, k)$ werden als orthogonal angenommen. D. h. es gilt

$$\mathbb{E}\{S(\mu, k)E_u(\mu, k)\} = 0. \quad (\text{A.20})$$

Wird außerdem angenommen, dass sich die Norm des Anregungssignals $\|\mathbf{X}(\mu, k)\|^2$ zwischen zwei aufeinanderfolgenden Blöcken nur wenig ändert, kann (A.19) vereinfacht werden zu

$$\alpha_{\text{opt}}(\mu, k) = \frac{\mathbb{E}\{|E_u(\mu, k)|^2\}}{\mathbb{E}\{|E(\mu, k)|^2\}}. \quad (\text{A.21})$$

Anhang B

Echtzeit-Rahmenwerk RACE

Um die im Rahmen dieser Arbeit entwickelten Signalverarbeitungsalgorithmen im Demonstratorfahrzeug testen zu können, wurde das Rahmenwerk RACE (engl. *Real-time Automotive Communication Enhancement*) entwickelt. Bei RACE handelt es sich um eine PC-basierte Software, innerhalb welcher die Signalverarbeitung implementiert ist. Dazu werden die im Fahrzeug verbauten Lautsprecher-Verstärker und Mikrofon-Vorverstärker über eine Audio-Soundkarte mit dem PC verbunden. Über eine grafische

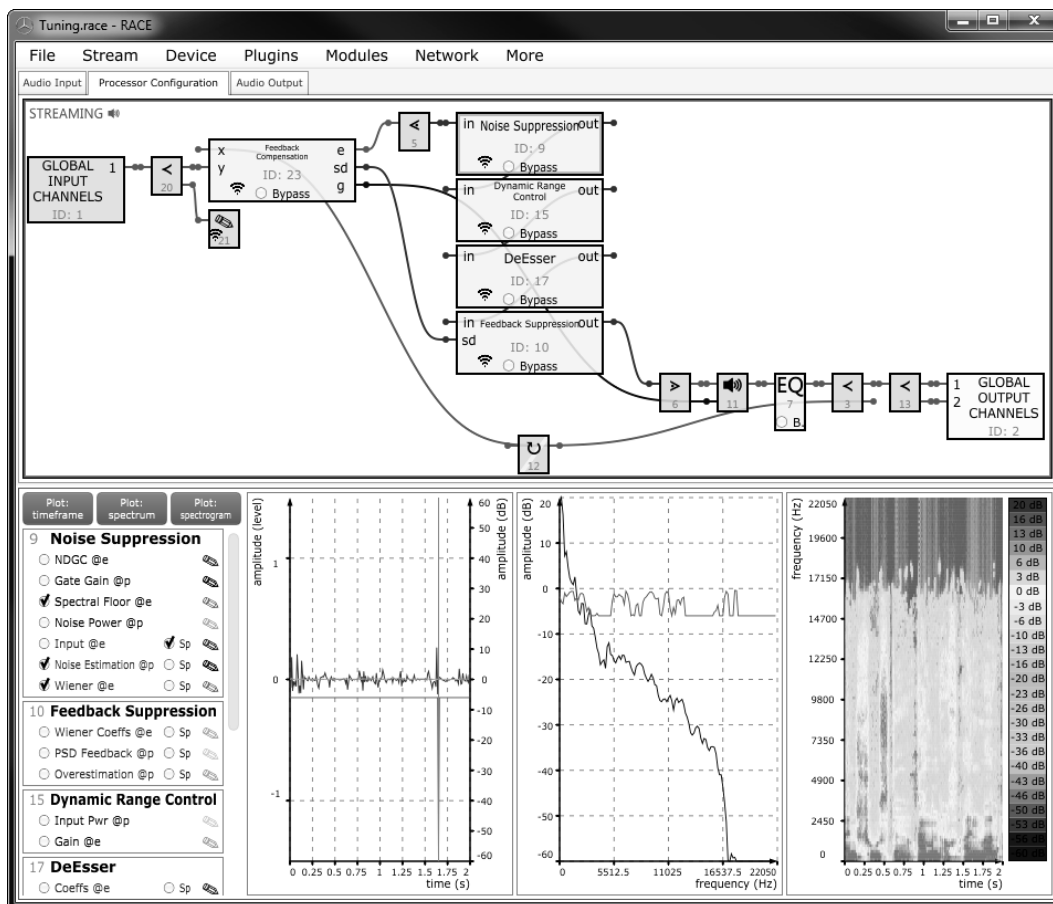


Abbildung B.1: Die grafische Benutzeroberfläche von RACE.

Benutzeroberfläche können die Algorithmen zur Laufzeit parametrieren und Signale in

Echtzeit visualisiert werden. Diese Benutzeroberfläche ist in Abb. B.1 dargestellt.

Die Software wurde mit der C++ Bibliothek JUCE [52] entwickelt, da diese sowohl die Funktionalität für die Audio-Schnittstelle als auch für die grafische Benutzeroberfläche bereitstellt. Für die Fouriertransformationen wurde die FFTW-Bibliothek [22] verwendet. Die Anbindung von Audio-Schnittstellen erfolgt mittels ASIO-Treiber, wodurch die Latenzzeit minimiert und mehrkanalige Audioverarbeitung ermöglicht wird. ASIO (engl. *Audio Stream Input/Output*) ist ein Audioprotokoll, welches von der Steinberg Media Technologies GmbH speziell für Echtzeit-Audioanwendungen entwickelt wurde [63].

Signalverarbeitungseinheiten werden in RACE in Form von RACE-*Audio-Modulen* eingebunden. Jedes Modul stellt zusätzlich zur Signalverarbeitung eine eigene grafische Benutzeroberfläche zur Verfügung, über welche es parametrierbar ist. Für das Modul zur Rückkopplungskompensation ist diese beispielhaft in Abb. B.2 gezeigt. Die

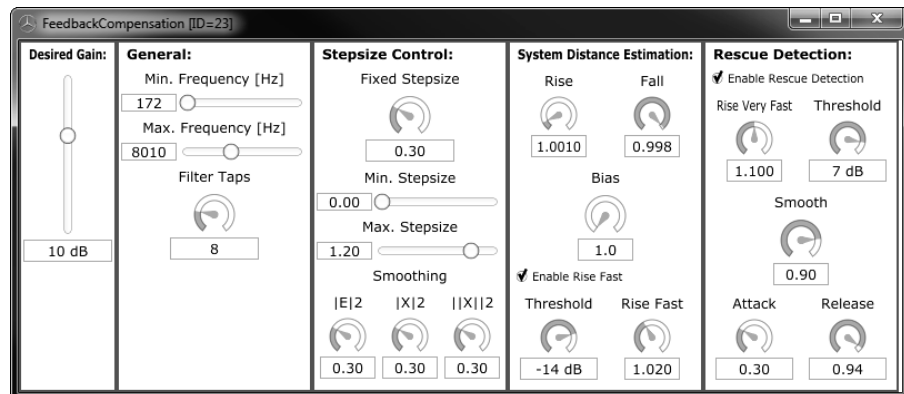


Abbildung B.2: Die grafische Benutzeroberfläche des Moduls für die Rückkopplungskompensation.

Module werden im Prozessor-Graph als Blöcke dargestellt. Ein Modul kann zur Laufzeit beliebig oft instanziiert werden. Die Ausgänge eines Moduls können grafisch über Leitungen mit den Eingängen der nachfolgenden Module verbunden werden. Dabei gibt es drei verschiedene Arten von Leitungen (vgl. Abb. B.1):

- Reellwertige Zeitbereichsdaten
- Komplexe Frequenzbereichsdaten
- Steuerdaten

Damit können die verschiedenen Module flexibel zu einer Signalverarbeitungseinheit kombiniert werden. Um zu gewährleisten, dass die zusammenschalteten Module in der korrekten Reihenfolge ausgeführt werden, wird in RACE eine automatische Topologiedetektion durchgeführt. Diese ermöglicht es, dass der Ausgang eines Moduls auf seinen eigenen Eingang zurückgekoppelt werden kann, was beispielsweise bei der Rückkopplungskompensation relevant ist.

Neue Module können über einen Codegenerator angelegt werden. Hierbei wird über eine Eingabemaske festgelegt, wie viele Eingangs-, Ausgangs- und Visualisierungskanäle das neue Modul benötigt. Außerdem können auch die Parameter festgelegt werden, welche über die Benutzeroberfläche einstellbar sein sollen. Der Codegenerator erzeugt darauf basierend die Quelltextdateien, innerhalb welcher die Signalverarbeitung sowie die grafische Benutzeroberfläche programmiert werden müssen. Die so erzeugten Quelltextdateien werden beim Kompilieren von RACE automatisch erkannt und als neues Modul integriert, ohne dass dazu am eigentlichen Rahmenwerk Änderungen vorgenommen werden müssen. So wird gewährleistet, dass verschiedene Entwickler neue Module ohne Vorkenntnis über das Rahmenwerk integrieren können.

Auf diese Weise wurden im Rahmen der vorliegenden Arbeit die folgenden Module in RACE implementiert:

- *Feedback Compensation* - Rückkopplungskompensator mit Nachhall-basierter Schrittweitensteuerung und Verstärkungsregelung, nach Kap. 4, bzw. [8].
- *Feedback Suppression* - Postfilter zur Unterdrückung der Restrückkopplungen, gesteuert über den geschätzten Systemabstand, nach Kap. 5, bzw. [24].
- *Noise Suppression* - Geräuschreduktion mit Wiener-Filter nach [43], geräuschabhängige Verstärkungsregelung und Noise Gate.
- *Exciter* - Der in Abschn. 6.2 bzw. [10] beschriebene Ansatz zur Erzeugung von Harmonischen für die Verbesserung der Sprachverständlichkeit.
- *De-Esser* - Der in [41] beschriebene Ansatz zur Verringerung von Zischlauten.
- *Frequency Shifter* - Ein Frequenzversatz zur Dekorrelation der Signale wie in [76] beschrieben.
- *Loss Control* - Pegelwaage zum adaptiven Umschalten zwischen zwei Kanälen.
- *Mic Mixer* - Mischer zum adaptiven Zusammenmischen zweier Mikrofonkanäle.
- *Equalizer* - Parametrischer Equalizer mit fünf Peak-Filtern und je einem Höhen- und Tiefen-Shelving-Filter nach [80].
- *Adaptive Notch Filter* - Adaptives Kerbfilter wie in [55] beschrieben.
- *Dynamic Range Control* - Modul zur Dynamikverarbeitung mittels Expander, Kompressor und Limiter nach [80].

Zusätzlich zu diesen RACE-Audio-Modulen können auch VST-Plugins in die Struktur eingebunden werden. VST (engl. *Virtual Studio Technology*) ist ein gängiges Plugin-Format in der Audiobranche und wurde ebenfalls von Steinberg entwickelt [63].

Für den Signalfluss gibt es sog. RACE *Build-In Tools*. Dazu gehören verschiedene Module mit denen Signale aufgeteilt oder zusammengeführt werden können, Signalgeneratoren, Verzögerungsglieder, etc. Die Soundkartenein- und -ausgänge sind ebenfalls

als Module dargestellt. Parallel zu den Soundkartenein- und -ausgängen können auch Audiodateien abgespielt bzw. aufgenommen werden.

Die RACE-Audio-Module können intern Signale für die Visualisierung zur Verfügung stellen. Darüber hinaus gibt es zusätzlich Senken für jeden Leitungstyp. Verbindet man eine Leitung mit der Senke, kann das Signal im Visualisierungsbereich gezeichnet werden. Im Visualisierungsbereich stehen drei verschiedene Arten zur Echtzeitvisualisierung der Signale zur Verfügung. Die Signale können hier entweder im Zeitbereich, als zeitabhängiges Spektrum oder als Spektrogramm angezeigt werden. Die Frequenzauflösung hängt hierbei von der Blockgröße der Audio-Soundkarte ab. Die Visualisierungen können als Bilddatei oder als Rohdaten abgespeichert werden, um sie auch außerhalb von RACE weiterverwenden zu können.

Über eine Netzwerkverbindung kann RACE mit einem Smartphone ferngesteuert werden. Dies ist insbesondere nützlich, um während Testfahrten nicht mit einem Laptop hantieren zu müssen. Über das Smartphone können beliebige Parameter der Module gesteuert werden. Darüber hinaus kann auch die Verarbeitung angehalten bzw. gestartet oder die Lautstärke an den Lautsprechern eingestellt werden. Für die Netzwerkkommunikation zwischen Smartphone und PC kommt das OSC-Protokoll (engl. *Open Sound Control*) zum Einsatz [20].

Mit RACE wurden im Rahmen der vorliegenden Arbeit zahlreiche verschiedene ICC Demonstrationen durchgeführt. Das Rahmenwerk hat sich dabei stets als zuverlässiges, unkompliziertes Werkzeug bewährt.

Literaturverzeichnis

- [1] Aphex. Exciter. <http://www.aphex.com/products/exciter>. Abgerufen am 06.04.2018.
- [2] Jacob Benesty, Hernán Rey, Leonardo Rey Vega und Sara Tressens. A Nonparametric VSS NLMS Algorithm. *IEEE Signal Processing Letters*, 13(10):581–584, 2006.
- [3] Elliott H. Berger, Larry H. Royster, Julia Doswell Royster, Dennis P. Driscoll und Martha Layne, Herausgeber. *The Noise Manual*, Band 5. American Industrial Hygiene Association, Fairfax, USA, 2000.
- [4] Steven F. Boll. Suppression of Acoustic Noise in Speech Using Spectral Subtraction. *IEEE Transactions on Acoustics Speech and Signal Processing*, ASSP-27(2):113–120, 1979.
- [5] Philipp Bulling, Florian Aue, Harald Schnepf und Arthur Wolf. Evaluierung eines Systems zur Verbesserung der Kommunikation zwischen Fahrer und Beifahrer in offenen Fahrzeugen. In *44. Deutsche Jahrestagung für Akustik (DAGA)*. München, Deutschland, 2018.
- [6] Philipp Bulling, Klaus Linhard, Arthur Wolf und Gerhard Schmidt. Acoustic Feedback Compensation with Reverb-based Stepsize Control for In-car Communication Systems. In *12th ITG Conference on Speech Communication*. Paderborn, Deutschland, 2016.
- [7] Philipp Bulling, Klaus Linhard, Arthur Wolf und Gerhard Schmidt. Approximation of the Optimum Stepsize for Acoustic Feedback Cancellation Based on the Detection of Reverberant Signal Periods. In *43. Deutsche Jahrestagung für Akustik (DAGA)*, Seiten 60–63. Kiel, Deutschland, 2017.
- [8] Philipp Bulling, Klaus Linhard, Arthur Wolf und Gerhard Schmidt. Stepsize Control for Acoustic Feedback Cancellation Based on the Detection of Reverberant Signal Periods and the Estimated System Distance. In *Conference of the International Speech Communication Association (INTERSPEECH)*. Stockholm, Schweden, 2017.
- [9] Philipp Bulling, Klaus Linhard, Arthur Wolf und Gerhard Schmidt. Automatic Equalization for In-car Communication Systems. In *Konferenz Elektronische Sprachsignalverarbeitung (ESSV)*. Ulm, Deutschland, 2018.

- [10] Philipp Bulling, Klaus Linhard, Arthur Wolf, Gerhard Schmidt, Anne Theiß und Marco Gimm. Nichtlineare Kennlinien zur Verbesserung der Sprachverständlichkeit in geräuschbehafteter Umgebung. In *42. Deutsche Jahrestagung für Akustik (DAGA)*. Aachen, Deutschland, 2016.
- [11] Josef Chalupper. Aural Exciter and Loudness Maximizer: What’s Psychoacoustic about “Psychoacoustic Processors”? In *109th Audio Engineering Society Convention*. Los Angeles, USA, 2000.
- [12] Jue Chang und John R. Glover. The Feedback Adaptive Line Enhancer: A Constrained IIR Adaptive Filter. *IEEE Transactions on Signal Processing*, 41(11):3161–3166, 1993.
- [13] Gerhard Doblinger. Computationally Efficient Speech Enhancement by Spectral Minima Tracking in Subbands. In *2nd European Conference on Speech Communication and Technology (EUROSPEECH)*. Madrid, Spanien, 1995.
- [14] Gerhard Doblinger. *MATLAB-Programmierung in der digitalen Signalverarbeitung*. J. Schlemmich Fachverlag, 2001.
- [15] Gerald Enzner und Peter Vary. Robust and Elegant, Purely Statistical Adaption of Acoustic Echo Canceller and Postfilter. In *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seiten 43–46. Kyoto, Japan, 2003.
- [16] Yariv Ephraim und David Malah. Speech Enhancement Using a Minimum Mean-Square Error Short-Time Amplitude Estimator. *IEEE Transactions on Acoustics Speech and Signal Processing*, ASSP-32(6):1109–1121, 1984.
- [17] Yariv Ephraim und David Malah. Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator. *IEEE Transactions on Acoustics Speech and Signal Processing*, ASSP-33(2):443–445, 1985.
- [18] Klaus Fellbaum. *Sprachverarbeitung und Sprachübertragung*. Springer-Verlag Berlin Heidelberg, 2. Auflage, 2012.
- [19] Jan Franzen und Tim Fingscheidt. A Delay-Flexible Stereo Acoustic Echo Cancellation for DFT-Based In-Car Communication (ICC) Systems. In *Conference of the International Speech Communication Association (INTERSPEECH)*. Stockholm, Schweden, 2017.
- [20] Adrian Freed und Matt Wright. Open Sound Control. <http://opensoundcontrol.org/>. Abgerufen am 06.04.2018.
- [21] Jürgen Freudenberger und Johannes Pittermann. Noise and Feedback Suppression for In-car Communication Systems. In *ITG-Fachtagung Sprachkommunikation*. Aachen, Deutschland, 2008.
- [22] Matteo Frigo und Steven G. Johnson. FFTW. <http://www.fftw.org/>. Abgerufen am 06.04.2018.

- [23] Pepe Gil-Cacho, Toon van Watershoot, Marc Moonen und Søren Holdt Jensen. Regularized Adaptive Notch Filters for Acoustic Howling Suppression. In *17th European Signal Processing Conference (EUSIPCO)*, Seiten 2574–2578. Glasgow, Schottland, 2009.
- [24] Marco Gimm, Philipp Bulling und Gerhard Schmidt. Energy Decay Based Post-filter for ICC Systems with Feedback Compensation. In *Konferenz Elektronische Sprachsignalverarbeitung (ESSV)*. Ulm, Deutschland, 2018.
- [25] Bastian Gröger, Michael Brodersen und Gerhard Schmidt. Verbesserung der Sprachverständlichkeit für Atemschutzmasken mittels Signalbearbeitung mit nichtlinearen Kennlinien. In *43. Deutsche Jahrestagung für Akustik (DAGA)*, Seiten 1477–1480. Kiel, Deutschland, 2017.
- [26] Meng Guo, Søren Holdt Jensen, Jesper Jensen und Steven L. Grant. On the Use of a Phase Modulation Method for Decorrelation in Acoustic Feedback Cancellation. In *20th European Signal Processing Conference (EUSIPCO)*, Seiten 2000–2004. Bukarest, Rumänien, 2012.
- [27] Emanuël A. P. Habets. Multi-Channel Speech Dereverberation Based on a Statistical Model of Late Reverberation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Band 4, Seiten iv/173–iv/176. Philadelphia, USA, 2005.
- [28] Eberhard Hänsler und Gerhard Schmidt. Hands-Free Telephones - Joint Control of Echo Cancellation and Postfiltering. *Signal Processing*, 80(11):2295–2305, 2000.
- [29] Eberhard Hänsler und Gerhard Schmidt. *Acoustic Echo and Noise Control - A Practical Approach*. John Wiley & Sons, Inc., 2004.
- [30] Tim Haulick, Gerhard Schmidt und Markus Buck. System for Equalizing an Acoustic Signal, 2012. United States Patent US 8,098,848 B2.
- [31] Simon Haykin. *Adaptive Filter Theory*. Prentice-Hall, Inc., 4. Auflage, 2002.
- [32] Ulrich Heute. Noise Reduction. In Eberhard Hänsler und Gerhard Schmidt, Herausgeber, *Topics in Acoustic Echo and Noise Control*. Springer, 2006.
- [33] International Telecommunication Union (ITU). ITU-T Recommendation P.800: Methods for Subjective Determination of Transmission Quality, 1996.
- [34] International Telecommunication Union (ITU). ITU-T Recommendation P.1110: Wideband Hands-Free Communication in Motor Vehicles, 2017.
- [35] Marc-André Jung und Tim Fingscheidt. A Shadow Filter Approach to a Wideband FDAF-Based Automotive Handsfree System. In *5th Biennial Workshop on DSP for In-Vehicle Systems*. Kiel, Deutschland, 2011.

-
- [36] Karl-Dirk Kammeyer und Kristian Kroschel. *Digitale Signalverarbeitung*. B. G. Teubner Verlag, Wiesbaden, 6. Auflage, 2006.
- [37] Heinrich Kuttruff. *Room Acoustics*. Taylor & Francis, 5. Auflage, 2009.
- [38] Rabea Landgraf, Johannes Köhler-Kaeß, Christian Lüke, Oliver Niebuhr und Gerhard Schmidt. Can you hear me now? Reducing the Lombard effect in a driving car using an In-Car Communication system. In *Speech Prosody*. Boston, USA, 2016.
- [39] Rabea Landgraf, Gerhard Schmidt, Johannes Köhler-Kaeß, Oliver Niebuhr und Tina John. More Noise, Less Talk - The Impact of Driving Noise and In-Car Communication Systems on Acoustic-Prosodic Parameters in Dialogue. In *43. Deutsche Jahrestagung für Akustik (DAGA)*. Kiel, Deutschland, 2017.
- [40] Katja Lebart, Jean-Marc Boucher und Philip N. Denbigh. A New Method Based on Spectral Subtraction for Speech Dereverberation. *Acta Acustica united with Acustica*, 87(3):359–366, 2001.
- [41] Klaus Linhard, Philipp Bulling und Arthur Wolf. Frequency Domain De-Essing for Hands-free Applications. In *43. Deutsche Jahrestagung für Akustik (DAGA)*. Kiel, Deutschland, 2017.
- [42] Klaus Linhard und Jürgen Freudenberger. Passenger In-Car Communication Enhancement. In *12th European Signal Processing Conference (EUSIPCO)*, Seiten 21–24. Wien, Österreich, 2004.
- [43] Klaus Linhard und Tim Haulick. Spectral Noise Subtraction with Recursive Gain Curves. In *5th International Conference on Spoken Language Processing (ICSLP)*. Sydney, Australien, 1998.
- [44] Klaus Linhard und Tim Haulick. Noise Subtraction with Parametric Recursive Gain Curves. In *6th European Conference on Speech Communication and Technology (EUROSPEECH)*. Budapest, Ungarn, 1999.
- [45] Eduardo Lleida, Enrique Masgrau und Alfonso Ortega. Acoustic Echo Control and Noise Reduction for Cabin Car Communication. In *7th European Conference on Speech Communication and Technology (EUROSPEECH)*, Seiten 1585–1588. Aalborg, Dänemark, 2001.
- [46] Andreas Mader, Henning Puder und Gerhard Schmidt. Step-size Control for Acoustic Echo Cancellation Filters - An Overview. *Signal Processing*, 80(9):1697–1719, 2000.
- [47] Jochen Marx. *Akustische Aspekte der Echokompensation in Freisprecheinrichtungen*. Dissertation, TU Darmstadt, Düsseldorf, Deutschland, 1996.
- [48] Harry Nyquist. Regeneration Theory. In *Bell System Tech. J.*, Band 11, Seiten 126–147. 1932.

- [49] Alfonso Ortega, Eduardo Lleida und Enrique Masgrau. Speech Reinforcement System for Car Cabin Communications. *IEEE Transactions on Speech and Audio Processing*, 13(5):917–929, 2005.
- [50] Alfonso Ortega, Eduardo Lleida, Enrique Masgrau und Fernando Gallego. Cabin Car Communication System to Improve Communications inside a Car. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Band 2. Orlando, Florida, 2002.
- [51] Henning Puder und Bernd Beigel. Controlling the Adaption of Feedback Cancellation Filters - Problem Analysis and Solution Approaches. In *12th European Signal Processing Conference (EUSIPCO)*, Seiten 25–28. Wien, Österreich, 2004.
- [52] ROLI Ltd. JUCE. <https://www.juce.com/>. Abgerufen am 06.04.2018.
- [53] Geert Rombouts, Toon van Watershoot und Marc Moonen. Robust and Efficient Implementation of the PEM-AFROW Algorithm for Acoustic Feedback Cancellation. *Journal of the Audio Engineering Society*, 55(11):955–966, 2007.
- [54] Marius Rotaru, Felix Albu und Henri Coanda. A Variable Step Size Modified Decorrelated NLMS Algorithm for Adaptive Feedback Cancellation in Hearing Aids. In *10th International Symposium on Electronics and Telecommunications (ISETC)*. Timisoara, Rumänien, 2012.
- [55] Gerhard Schmidt und Tim Haulick. Signal Processing for In-Car Communication Systems. In E. Hänsler und G. Schmidt, Herausgeber, *Topics in Acoustic Echo and Noise Control*, Kapitel 14, Seiten 437–493. Springer, Berlin, 2006.
- [56] Gerhard Schmidt, Tim Haulick und Markus Buck. Equalization in Acoustic Signal Processing, 2007. European Patent Application EP 1 858 295 A1.
- [57] Gerhard Schmidt, Anne Theiß, Jochen Withopf und Arthur Wolf. Evaluation of In-car Communication Systems. In John H.L. Hansen, Pinar Boyraz, Kazuya Takeda und Hüseyin Abut, Herausgeber, *Digital Signal Processing for In-Vehicle Systems and Safety*, Seiten 73–107. Springer New York, 2012.
- [58] Manfred R. Schroeder. Improvement of Acoustic-Feedback Stability by Frequency Shifting. *The Journal of the Acoustical Society of America*, 36(9):1718–1724, 1964.
- [59] Christian Schüldt, Frederic Lindstrom, Haibo Li und Ingvar Claesson. Adaptive Filter Length Selection for Acoustic Echo Cancellation. *Signal Processing*, 89(6):1185–1194, 2009.
- [60] Priyanka Shekar und Julius Smith. Modeling the Harmonic Exciter. In *135th Audio Engineering Society Convention*. New York, USA, 2013.
- [61] John J. Shynk. Frequency-Domain and Multirate Adaptive Filtering. *IEEE Signal Processing Magazine*, 9(1):14–37, 1992.

-
- [62] Jia-Sien Soo und Khee K. Pang. Multidelay Block Frequency Domain Adaptive Filter. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(2):373–376, 1990.
- [63] Steinberg Media Technologies GmbH. ASIO SDK und VST SDK. <https://www.steinberg.net/de/company/developer.html>. Abgerufen am 06.04.2018.
- [64] Falco Strasser und Henning Puder. Sub-band Feedback Cancellation with Variable Step Sizes for Music Signals in Hearing Aids. In *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, Seiten 8207–8211. Florenz, Italien, 2014.
- [65] Falco Strasser und Henning Puder. Adaptive Feedback Cancellation for Realistic Hearing Aid Applications. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 23(12):2322–2333, 2015.
- [66] Falco Strasser und Henning Puder. Correlation Detection for Adaptive Feedback Cancellation in Hearing Aids. *IEEE Signal Processing Letters*, 23(7):979–983, 2016.
- [67] Linh T. T. Tran, Henning Schepker, Simon Doclo, Hai H. Dam und Sven Nordholm. Improved Practical Variable Step-Size Algorithm For Adaptive Feedback Control in Hearing Aids. In *10th International Conference on Signal Processing and Communication Systems (ICSPCS)*. Gold Coast, Australien, 2016.
- [68] Toon van Watershoot und Marc Moonen. Adaptive Feedback Cancellation for Audio Applications. *Signal Processing*, 89(11):2185–2201, 2009.
- [69] Toon van Watershoot und Marc Moonen. Comparative Evaluation of Howling Detection Criteria in Notch-Filter-based Howling Suppression. In *126th Convention of the Audio Engineering Society*. München, Deutschland, 2009.
- [70] Toon van Watershoot und Marc Moonen. Fifty Years of Acoustic Feedback Control: State of the Art and Future Challenges. *Proceedings of the IEEE*, 99(2):288–327, 2011.
- [71] Peter D. Welch. The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. *IEEE Transactions on Audio and Electroacoustics*, AU-15(2):70–73, 1967.
- [72] Bernard Widrow und Marcian Edward Hoff. Adaptive Switching Circuits. In *Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record*, Band 4, Seiten 96–104. Los Angeles, USA, 1960.
- [73] Bernard Widrow und Samuel D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, Inc., Upper Saddle River, New Jersey, USA, 1985.
- [74] Jochen Withopf. *Signalverarbeitungsverfahren zur Verbesserung der Sprachkommunikation im Fahrzeug*. Dissertation, Christian-Albrechts-Universität zu Kiel, 2017.

- [75] Jochen Withopf, Laila Jassoume, Gerhard Schmidt und Anne Theiß. A Modified Overlap-Add Filter Bank With Reduced Delay. In *38. Deutsche Jahrestagung für Akustik (DAGA)*. Darmstadt, Deutschland, 2012.
- [76] Jochen Withopf, Sebastian Rhode und Gerhard Schmidt. Application of Frequency Shifting in In-Car Communication Systems. In *11th ITG Conference on Speech Communication*. Erlangen, Deutschland, 2014.
- [77] Jochen Withopf und Gerhard Schmidt. Estimation of Time-Variant Acoustic Feedback Paths in In-Car Communication Systems. In *14th International Workshop on Acoustic Signal Enhancement (IWAENC)*. Antibes, Frankreich, 2014.
- [78] Arthur Wolf und Bernd Iser. Energy Decay Model Based Feedback Suppression: Theory and Application. In *5th Biennial Workshop on DSP for In-Vehicle Systems*. Kiel, Deutschland, 2011.
- [79] Arthur Wolf, Bernd Iser und Gerhard Schmidt. Laufzeitoptimierte Geräuschreduktionsverfahren basierend auf Overlap-Save-Strukturen mit Projektionsfilternäherungen. In *Konferenz Elektronische Sprachsignalverarbeitung (ESSV)*. Berlin, Deutschland, 2010.
- [80] Udo Zölzer. *Digitale Audiosignalverarbeitung*. Vieweg+Teubner, GWV Fachverlage GmbH, Wiesbaden, Deutschland, 3. Auflage, 2005.