UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

FAKULTÄT FÜR
ELEKTROTECHNIK,
INFORMATIK UND
MATHEMATIK

Mehrkanalige Sprachsignalverbesserung durch adaptive Lösung eines Eigenwertproblems im Frequenzbereich

Zur Erlangung des akademischen Grades

DOKTORINGENIEUR (Dr.-Ing.)

der Fakultät für Elektrotechnik, Informatik und Mathematik
der Universität Paderborn
vorgelegte Dissertation
von

Dipl.-Ing. Ernst Warsitz
Oppeln

Referent: Prof. Dr.-Ing. Reinhold Häb-Umbach
Korreferent: Prof. Dr.-Ing. Peter Vary

Tag der mündlichen Prüfung: 12.12.2008

Paderborn, den 03.03.2009

Diss. EIM-E/248

Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter im Fachgebiet Nachrichtentechnik der Universität Paderborn. Insbesondere die spannende Aufbauphase beim Etablieren der neuen Schwerpunkte Sprachsignalverarbeitung und Spracherkennung während der Anfangszeit meiner Tätigkeit gemeinsam mit dem Fachgebietsleiter Herrn Prof. Dr.-Ing. Reinhold Häb-Umbach werden mir im Gedächtnis bleiben. Ihm danke ich für die sehr gute Arbeitsatmosphäre, für die zahlreichen anregenden Diskussionen und für die Übernahme des Referates dieser Arbeit. Herrn Prof. Dr.-Ing. Peter Vary von der Universität Aachen danke ich für die Übernahme des Korreferates und die Hinweise zur Verbesserung dieser Arbeit.

Im Zuge meiner Tätigkeit sind eine Reihe von Projekt-, Studien- und Diplomarbeiten entstanden, deren Ergebnisse vielfältig in die Dissertation eingeflossen sind. Allen Studenten danke ich für die gute Zusammenarbeit. Stellvertretend seien hier Herr Dipl.-Ing. Maik Bevermeier, Herr Dipl.-Math. Alexander Krüger, Herr Dipl.-Ing. Jörg Schmalenströer und Herr Dipl.-Ing. Dang Hai Tran Vu erwähnt, die mir nach ihrer studentischen Tätigkeit als Kollegen erhalten geblieben sind. Ihnen und meinen weiteren Kollegen danke ich für die vielen fachlichen und freundschaftlichen Gespräche. Meinem Kollegen Herrn Dipl.-Inf. Sven Peschke danke ich insbesondere für die anregende Zeit im gemeinsamen Büro und den unkonventionellen fachlichen Gedankenaustausch. Für die hervorragende Unterstützung bei der fachgebietseigenen Simulationssoftware und die liebenswerten Kommentare zu allen Lebenslagen danke ich Herrn Dr.-Ing. Valentin Ion.

Meinen Seilpartnern Georg, Jörn und Ingo danke ich für die schönen Stunden in der Natur und in der Vertikalen. Sie haben mir mit dem Klettern eine ideale Abwechslung zum Uni-Alltag ermöglicht und mir geholfen, die aufreibenden Phasen während der Promotionszeit durchzustehen.

Meiner Frau Kerstin danke ich für ihre unglaubliche Geduld, den Verzicht auf viele gemeinsame Wochenenden und das Ertragen angespannter Arbeitsphasen. Durch ihre tatkräftige Unterstützung im Alltag hat sie mir eine intensive Auseinandersetzung mit dieser Arbeit ermöglicht. Meiner Tochter Frieda danke ich für ihr Lachen und ihre unendliche Liebe.

Abschließend gilt mein Dank den Menschen, die mich von erster Stunde an begleitet haben, meinen Eltern. Ihr Vertrauen, ihre Großzügigkeit und ihre stetige Unterstützung haben für mich erst viele Wege in meinem Leben gangbar gemacht – so auch das Studium und die Promotion.

Inhaltsverzeichnis

Kapitel 1

Einleitung

Die wichtigste und natürlichste Kommunikationsform des Menschen ist die Sprache. Aufgrund der Entwicklung der Informations- und Kommunikationstechnik in den letzten Jahren sind viele Anwendungen entstanden, um dem Bedürfnis des Menschen nach allgegenwärtiger und komfortabler Sprachkommunikation zu entsprechen. Dies ist insbesondere an dem sprunghaft gewachsenen Markt der Mobiltelefonie im letzten Jahrzehnt abzulesen. Zusätzlich zu der mobilen Telefonie und der klassischen Festnetztelefonie entsteht aktuell ein besonderes Interesse an neuen Kommunikationstechniken wie sie die internetbasierte Telefonie ermöglicht. Dabei ist insbesondere der Vorteil zur Sprachkommunikation, parallel weitere Daten wie Text-, Bild- und Videomaterial auszutauschen, sehr reizvoll.

Um den Komfort für Kommunikationsteilnehmer zu steigern und eine erhöhte Mobilität während der Kommunikation zu gewährleisten, ist der Einsatz von Freisprecheinrichtungen wünschenswert[1]. Im überwiegenden Fall ist dabei der Kommunikationspartner ebenfalls ein Mensch, weshalb auch von Mensch-Mensch-Kommunikation gesprochen wird. Dadurch offenbaren sich aber auch schon zwei wesentliche Probleme der Freisprechanwendung. Zum einen werden bei der Aufnahme der Sprache vorhandene Störgeräusche in der Umgebung des Sprechers ebenfalls von den Mikrophonen erfasst und mitübertragen. Zum anderen entstehen bei einer Duplex-Verbindung durch die gleichzeitige Ausgabe und Aufnahme der Sprache beim entfernten Kommunikationspartner auf der Sendeseite Echo- bzw. Halleffekte. Die Einbußen in der Sprachqualität durch die additiven Störgeräusche und die äußerst irritierenden Echokomponenten müssen durch geeignete Verfahren zur Sprachsignalverbesserung minimiert werden, um eine Akzeptanz des Anwenders für ein Freisprechsystem zu schaffen.

Während bei einer Mensch-Mensch-Kommunikation Störungen der Sprache lediglich als qualitätsmindernd anzusehen sind, können gestörte Sprachsignale bei der maschinellen Weiterverarbeitung zu erheblichen Fehlern führen. Bei dieser so genannten Mensch-Maschine-Kommunikation ist eine hohe Sprachsignalqualität von essentieller Bedeutung für das maschinelle Erkennen der gesprochenen Sprache. Die automatische Spracherkennung kann dabei z. B. in Auskunftsystemen, für Diktieranwendungen oder aber auch zur Sprachsteuerung von Robotersystemen eingesetzt werden. Verwandt mit der Spracherkennung ist die Sprechererkennung. Diese kann zur Stimmenidentifizierung und Authentifizierung in Sicherheitssystemen zum Einsatz kommen. Weiterhin kann z. B. für Überwachungssysteme, oder zur

[1] Aus Gründen der Verkehrssicherheit ist das Telefonieren im fahrenden Kraftfahrzeug ohne Freisprecheinrichtung seit Februar 2001 sogar verboten.

Kamerasteuerung bei einer Telekonferenz die Information über die Position des Sprechers in seiner Umgebung von Bedeutung sein. All diese Problemstellungen können auch als Teilaspekte eines Gesamtsystems verstanden werden, welches dann mit dem Schlagwort akustische Szenenanalyse bezeichnet werden kann. Während aber bei der akustischen Szenenanalyse alle präsenten Geräuschquellen zu lokalisieren, zu trennen und letztlich zu klassifizieren sind, ergibt sich unter dem Gesichtspunkt der Sprache als Quellsignal vielmehr die griffig zu formulierende Fragestellung: "Wer spricht wann, wo und was?"

Die genannten Anwendungsbeispiele zeigen die hohe Relevanz der Sprachsignalverarbeitung und insbesondere den Teilaspekt der Sprachsignalverbesserung für gegenwärtig existierende Produktlösungen aber auch für zukunftsweisende Szenarien wie z. B. eine vernetzte Hausumgebung mit intelligenten, multimodalen Schnittstellen als kontextbewusstes System [Ami, SHUW07]. Hierbei soll die Problemstellung der Sprachsignalverbesserung unter dem Aspekt lediglich eines aktiven Sprechers[2] in einer gestörten Umgebung betrachtet werden. Bezüglich der Störquellen soll die praktisch sehr relevante Annahme gelten, dass es sich um ein schwach zeitvariantes Hintergrundrauschen mit einer leichten Tiefpasscharakteristik handelt. Grundsätzlich lassen sich dabei Geräuschreduktionssysteme in zwei Kategorien einteilen: ein- und mehrkanalige Systeme, d. h. ob die zugrundeliegende Freisprecheinrichtung aus einem oder mehreren Mikrophonen aufgebaut ist. Im Bereich der einkanaligen Sprachsignalverbesserung sind in den letzten Jahrzehnten eine Vielzahl von Verfahren vorgestellt worden, die im Wesentlichen rein spektrale Eigenschaften der Eingangsdaten auswerten, siehe z. B. [BCM05, VM06, Loi07]. In der Regel ist mit einkanaligen Methoden zwar eine starke Unterdrückung der Störsignalkomponenten möglich, aber gleichzeitig entstehen zusätzliche Artefakte als tonales Rauschen[3] und eine nicht unerhebliche Sprachverzerrung.

Durch die Anordnung von mehreren akustischen Sensoren als Mikrophongruppe entsteht ein mehrkanaliges Signal zur Weiterverarbeitung mittels Algorithmen zur Störgeräuschreduktion. Die mehrkanaligen Ansätze zur Signalverarbeitung können dann zusätzlich die räumliche Information, welche aufgrund von Laufzeitunterschieden der akustischen Signale von den Quellen zu den Mikrophonen entstehen, ausnutzen. Dadurch ist eine gewisse räumliche Unterscheidung zwischen dem Nutzsignal und dem Störsignal möglich, wobei Signale aus der Richtung des Nutzsignals möglichst unverzerrt zu belassen und Störsignale aus anderen Raumrichtungen zu unterdrücken sind. Diese Vorgehensweise kann auch als akustische Strahlformung[4] (engl. *Beamforming*) aufgefasst und über die so genannte Richtcharakteristik räumlich interpretiert werden.

In praktischen Anwendungsbeispielen treten vielfältige Probleme beim Entwurf von *Beamforming*-Verfahren auf, wie der breitbandige Charakter der Sprachsignale, die Mehrwegeausbreitung des akustischen Schalls aufgrund von Reflexionen an den Raumbegrenzungen, und mögliche Positionsänderungen der akustischen Quellen, insbesondere ein sich bewegender Sprecher. Außerdem unterliegt das Design wirtschaftlichen und geometrischen Restriktionen bezüglich der Anzahl und der Anordnung der Mikrophone.

[2] Die Problemstellung, ein Gemisch von mehreren Sprachsignalen zu trennen, wird nur peripher im Anhang F behandelt.

[3] Bei dem tonalen Rauschen handelt es sich um spektral gefärbtes, instationäres Rauschen, welches für den Menschen als sehr unangenehm empfunden wird.

[4] Eine räumliche Filterung von Signalen wird in verschiedenen Anwendungen eingesetzt wie z. B. der Radartechnik, der Kommunikationstechnik, bei geophysikalischen Anwendungen oder eben auch in der Sprachsignalverarbeitung.

Seit einigen Jahren wird die Verwendung von Mikrophongruppen zur Sprachsignalverarbeitung immer interessanter aufgrund der stetig steigenden Leistungsfähigkeit von digitalen Signalprozessoren.

1.1 Mehrkanalige Störgeräuschreduktion

Die entscheidende Eigenschaft eines mehrkanaligen Mikrophonsignals besteht im Zeitversatz der einzelnen Signale zueinander bedingt durch die Laufzeitunt erschiede des akustischen Signals von der Quelle hin zu den verwendeten Sensoren. Dies gilt für die direkten Ausbreitungspfade aber auch für die jeweiligen Reflexionen an den Raumbegrenzungen. Die einfachste Variante einer Strahlformung besteht nun darin, den entstandenen Zeitversatz der direkten Ausbreitungspfade zu kompensieren und die Signale kohärent zu addieren. Ein solches Verfahren wird entsprechend dem Vorgehen als *Delay-and-Sum-Beamformer* (DSB) bezeichnet. Dabei wird das akustische Signal aus einer gewünschten Richtung unverzerrt übertragen, wohingegen Signale aus anderen Richtungen gedämpft werden. Zusätzlich ist es möglich, nach dem Laufzeitausgleich in jedem Signalpfad Filterfunktionen einzusetzen, welche je nach Entwurfskriterium die Richtcharakteristik des *Beamformers* beeinflussen. Allgemein kann dann von einem *Filter-and-Sum-Beamformer* (FSB) gesprochen werden. Dabei lässt sich der Entwurf für die Filter prinzipiell in zwei Klassen aufteilen: datenunabhängige und datenabhängige Verfahren [VVB88]. Bei den datenunabhängigen Verfahren werden die Filtergewichte a priori entsprechend einer gewünschten Richtcharakteristik entworfen [VT02] und bleiben während der Anwendung unverändert; sie sind also unabhängig von den Eigenschaften der zu verarbeitenden Signale. Im Allgemeinen ist jedoch eine deutlich höhere Unterdrückung von Störgeräuschen zu erzielen, wenn die Filterkoeffizienten des *Beamformers* adaptiv dem konkreten Störszenario angepasst werden können. In diesem Fall spricht man von datenabhängigen *Beamforming*-Verfahren.

Grundsätzlich bedeutet mehrkanalige Störgeräuschreduktion aus Sicht der statistischen Signalverarbeitung die Minimierung der Varianz der Störung am Ausgang des *Beamformers*. Dieses Ziel wird je nach Anordnung der Mikrophone, sowie a priori Wissen und Annahmen bezüglich der Signale mit unterschiedlichen Ansätzen der Signalverarbeitung verfolgt. Eine sehr übersichtliche Einführung in die Thematik des adaptiven *Beamformings* präsentieren Van Veen und Buckley in [VVB88]. Cox et al. haben eine Zusammenfassung früher Arbeiten zum adaptiven *Beamforming* und zu Robustheitsaspekten in [CZO87] erstellt. Das wohl bekannteste Optimierungskriterium besteht in der Minimierung der Ausgangsleistung des *Beamformers* mit der Nebenbedingung eines unverzerrten Signals aus einer gewünschten Richtung, welche im Allgemeinen durch die Schätzung der Sprecherrichtung gegeben ist. Daher wird auch vom *Minimum Variance Distortionless Response* (MVDR) *Beamformer* gesprochen. Für den MVDR *Beamformer* sind in [GM55] und [Cap69] erste Untersuchungen von besonderer Bedeutung zu finden, wobei der Fokus auf dem seismischen Anwendungsbereich liegt. Einen weit verbreiteten adaptiven MVDR-Lösungsansatz für das akustische *Beamforming* hat Frost [Fro72] vorgestellt. Dabei wird für die geschätzte Sprecherrichtung eine vorgegebene spektrale Übertragungsfunktion mittels einer Nebenbedingung[5] eingehalten, während die Leistung des

[5] Da es sich in der Regel um lineare Nebenbedingungen beim adaptiven MVDR *Beamformer* handelt, wird auch häufig vom *Linearly Constrained Minimum Variance Distortionless Response* (LCMVDR) *Beamformer* gesprochen.

Rauschens durch die Minimierung der gesamten Ausgangsleistung reduziert wird. Während die Gradienten-Adaptionsregel in [Fro72] mit den instantanen Eingangsdaten abläuft, erfolgt in [Dob06] basierend auf der Arbeit in [PK01] dessen Erweiterung unter Berücksichtigung der geglätteten spektralen Kreuzleistungsdichten der Signale mit dem Ziel einer beschleunigten Adaption. Eine theoretische Basis für die MVDR-Lösung unter Einbeziehung der Mehrwegeausbreitung von akustischen Signalen wird in [KHJ06] vorgestellt[6].

Die Leistungsfähigkeit von MVDR-*Beamforming*-Verfahren wird entscheidend von der Genauigkeit der Schätzung der Einfallsrichtung des gewünschten Quellensignals, und damit der Richtung aus der das Signal unverzerrt übertragen werden soll, bestimmt. Abweichungen zwischen wahrer und geschätzter Richtung, können zu starken Signalverzerrungen und ungewollter Verstärkung von Störungen führen [WA96]. Weiterhin sind die meisten *Beamforming*-Verfahren sehr sensitiv gegenüber nicht kalibrierten Mikrophonsystemen (ungleiche Richtcharakteristiken der verwendeten Mikrophone, unterschiedliche Verstärkung der einzelnen Signalpfade in der nachverarbeitenden Elektronik und ungenaue Positionierung der Mikrophone). Daher beschäftigt sich eine Vielzahl aktueller Arbeiten zu adaptiven MVDR-*Beamforming*-Methoden mit Robustheitsaspekten. Hierzu sind in [LS05] wesentliche Methoden beschrieben und ebenfalls in [HGJ06, JHLCCC06] erwähnt.

Wird ein *Beamformer* ausschließlich hinsichtlich der Direktivität optimiert, also das Signal-zu-Rauschleistungsverhältnis für den Fall eines diffusen Störschallfelds maximiert, so erhält man eine spezielle Klasse von MVDR-*Beamforming*-Verfahren, die in der Praxis von großer Bedeutung sind. Man bezeichnet solch einen *Beamformer* als Superdirektiven *Beamformer* und dessen Eigenschaft als Superdirektivität. Frühe Arbeiten zur robusten Realisierung Superdirektiver *Beamformer* sind in [GM55, US56, CZK86] und eine aktuellere Übersicht in [Elk00, BS01] zu finden. Moderne Realisierungen beinhalten z. B. Entwurfskriterien für Nahfeldanwendungen wie in [Täg98, JG00] mit der Anwendung für Freisprecheinrichtungen in Kraftfahrzeugen [MPL01] oder zur Spracherkennung an einem PC-Arbeitsplatz [MMM00]. Weitere aktuelle Arbeiten beschäftigen sich z. B. mit Robustheitsaspekten bezüglich fehlerhafter Annahmen der Charakteristik von linearen Mikrophongruppen kleiner Apertur [DM06], dem *Beamforming*-Design für binaurale Anwendungen mit zweikanaliger Ein- und Ausgabe zur Bewahrung Interauraler Eigenschaften [LV06] oder der Einbeziehung der Richteigenschaften der verwendeten Mikrophone [Buc07]. Bitzer et al. stellen in [BSK99a] eine alternative Realisierung des Superdirektiven *Beamformers* in einer Struktur als *Generalized Sidelobe Canceller* (GSC) mit dem Vorteil einer Reduzierung der Rechenkomplexität vor.

Grundsätzlich erfolgt bei einem GSC die Minimierung des Rauschens in einem Signal, welches mit einem nichtadaptiven *Beamformer* erzeugt wird, mittels adaptiver Filter, an dessen Eingängen dann Störgeräuschreferenzsignale anliegen. Diese Störgeräuschreferenzsignale werden mit Hilfe einer so genannten *Blocking Matrix* erzeugt. Die GSC-Struktur wurde erstmals von Griffiths und Jim [GJ82] vorgeschlagen und kann als Umformung des bedingten Minimierungsproblems nach [Fro72] in ein Minimierungsproblem ohne Nebenbedingung betrachtet werden. In [GJ82] wird vorgeschlagen, die Störgeräuschreferenzsignale durch die paarweise Subtraktion aufeinander zeitangepasster Signale zu generieren.

Bitzer et al. [BSK99c] sowie Nordholm und Leug [NL00] haben den GSC abhängig von dem Störschallfeld untersucht. Für den Fall von gerichteten Störungen ist dabei die Rausch-

[6] Obschon bei den Herleitungen in [KHJ06] Reflexionspfade berücksichtigt werden, so ist in den Experimenten nur der direkte Ausbreitungspfad zu finden.

unterdrückung theoretisch unendlich hoch, während bei dem praktisch sehr relevanten diffusen Störschallfeld die Geräuschreduktion recht gering ausfällt. Ein wesentliches Problem der Originalvariante nach [GJ82] ist die Annahme der Freifeldausbreitung des Sprachsignals. Denn nur unter dieser Bedingung können mittels der paarweisen Subtraktion aufeinander zeitangepasster Mikrophonsignale optimale Störgeräuschreferenzsignale erzeugt und eine hohe Störgeräuschreduktion bei unverzerrt gebliebenem Sprachsignal erreicht werden. Dieses Manko ist in einigen Arbeiten explizit aufgegriffen worden.

Nordholm et al. haben in [NCB93] räumliche Hochpassfilter in der *Blocking Matrix* verwendet. Durch die aufwendige Filterung sind dann genauere Störgeräuschreferenzsignale bestimmt worden. Meyer und Sydow [MS97] verwenden unterschiedliche *Beamformer* für die Störung und den Sprecher, um mittels des *Beamformers* für das Störsignal den Anteil der Sprache im Störgeräuschreferenzsignal zu vermindern.

Die Mehrwegeausbreitung des Sprachsignals ist von Jan und Flanagan in [JF96] konstruktiv mittels *Matched Filter* im nichtadaptiven *Beamformer* genutzt worden. Die Filter bestehen dabei aus komplex konjugierten Übertragungsfunktionen, welche zuvor zwischen dem Sprecher und den Sensoren bestimmt wurden. Rabinkin et al. [RRFM98] zeigen, dass solch ein *Matched Filter Beamformer* (MFB) einem DSB überlegen ist.

Eine adaptive Variante beschreiben Gazor et al. [GAG96], wobei das Nachführen der Filter durch eine iterative Hauptkomponentenanalyse der spektralen Kreuzleistungsdichtematrix der Eingangsdaten mittels einer modifizierten Variante des Adaptionsverfahrens [Oja82] erfolgt. Dabei wird das Ausgangssignal des GSCs zur Adaption des *Matched Filter Beamformers* rückgekoppelt. Entscheidend für die Adaption sind die Initialwerte der Filterkoeffizienten. Die *Blocking Matrix* wird äquivalent zu [GJ82] berechnet. Eine Erweiterung dieses Verfahrens ist in [AG97] zu finden mit der Adaptionsregel [Yan95] und einem expliziten Lösungsvorschlag zur Normalisierung der *Matched-Filter*-Koeffizienten optimiert für eine Sprecherposition vor einem PC-Arbeitsplatz. Hier wird die *Blocking Matrix* zur Erzeugung der Störgeräuschreferenzsignale durch eine orthogonale Projektion bestimmt. Bei der Anwendung in einer Umgebung mit unbekannten, gerichteten Störschallquellen kann jedoch eine ungewollte Identifizierung der Störung als Nutzsignal vorkommen und vice versa das Sprachsignal unterdrückt werden.

Hoshuyama et al. [HSH96, HSH99] haben *Least Mean Squares* (LMS) adaptive Filter zur Sprachsignalunterdrückung in der *Blocking Matrix* verwendet und benutzen so genannte leckende (engl. *Leaky*) Koeffizienten bzw. eine Koeffizientenbeschränkung zur Robustheitssteigerung bezüglich einer fehlerhaften Sprecherrichtungsschätzung. In Phasen, wenn am Ausgang des nichtadaptiven *Beamformers* lediglich das Sprachsignal beobachtet wird, dient dieses dann als Referenz zur Adaption der *Blocking Matrix*. In einem Szenario mit permanent aktiven Störschallquellen, sind solche Zeitabschnitte jedoch nicht vorhanden. Die Adaption mit einem stark gestörten Sprachsignal führt dann konsequenterweise zu erheblichen Sprachsignalverzerrungen durch den GSC. Die Struktur des GSCs mit LMS-adaptiver *Blocking Matrix* und LMS-adaptiven Filtern zur Rauschunterdrückung ist als effiziente Realisierung im Frequenzbereich von Herbordt und Kellermann in [HK01] vorgestellt worden. Die resultierende GSC-Struktur wurde in [Her04] mit einer Echokompensation in unterschiedlichen Varianten als Gesamtsystem realisiert und untersucht. In [HBNK07] sind weitere Robustheitsaspekte bezüglich der Adaption beschrieben, um Probleme bedingt durch das so genannte Gegensprechen in Freisprecheinrichtungen zu lösen.

Eine signalangepasste *Blocking Matrix*, welche auch mit einem stark gestörten Sprachsi-

gnal adaptiert werden kann, wurde von Gannot et al. [GBW99, GBW01] vorgestellt. Grundlage ist dabei die Schätzung der Verhältnisse der Raumübertragungsfunktionen zwischen dem Sprecher und den Mikrophonen nach dem in [SW96] beschriebenen Kriterium der Dekorrelation unter Ausnutzung der Stationarität des Störsignals und der Nichtstationarität der Sprache. Die entstehenden Sprachverzerrungen des Gesamtsystems sind ausführlich in [GBW04] behandelt. Dabei scheinen insbesondere in dem unteren Frequenzbereich Probleme aufzutreten. Die GSC-Struktur ist zur weiteren Störgeräuschreduktion in [GC04] mit einer zusätzlichen Nachfilterung versehen worden. Eine Erweiterung des Gesamtsystems zur Unterdrückung eines zweiten Sprechers – also einer instationären Störquelle – wird in [RGC07b, RGC08] vorgestellt.

Eine andere Variante des adaptiven *Beamformings* ergibt sich mit dem Ansatz der Minimierung des kleinsten mittleren quadratischen Fehlers (engl. *Minimum Mean Squared Error*, MMSE). Dabei besteht die Schwierigkeit in der Schätzung eines Referenzsignals. Der populärste Ansatz hierbei ergibt sich in der sequenziellen Anordnung eines MVDR *Beamformers* und eines einkanaligen Nachfilters (engl. *Postfilter*) [SBM01], wobei eine Mittelung der Kreuzleistungsdichten zwischen jeweils zwei Signalpaaren zur Schätzung der spektralen Kreuzleistungsdichte-Matrix des Nutzsignals verwendet werden kann [Zel88]. Eine Verbesserung dieser Schätzung ist Gegenstand neuerer Veröffentlichungen [SW92, MMU98, BSK99b, MB02, MB03].

Alternativ wurden von Nordholm et al. [NCG01] über eine Kalibrierungs-Sprachsequenz die optimalen Filterkoeffizienten für die Mikrophongruppe (*In Situ Calibrated Microphone Array*, ICMA) in einem Kraftfahrzeug berechnet und eine Teilbandimplementierung vorgenommen. Dabei beinhaltet die MMSE-Schätzung repräsentative Einflüsse der verwendeten Hardware sowie der Mikrophon- und Sprecherposition. In [GN02, NGL05] ist dieser Ansatz für eine gewisse Region (*Soft Constrained*) um die erwartete Sprecherrichtung erweitert.

Eine andere Möglichkeit zur Schätzung eines Referenzsignals basiert auf Techniken ähnlich denen zur einkanaligen spektralen Subtraktion. Dafür wird in [Flo01] eine Sprache/Pause-Detektion eingesetzt und die *Beamformer*-Adaption über einen LMS-Algorithmus durchgeführt. Eine Adaption nach dem RLS -Prinzip (engl. *Recursive Least Squares*, RLS) gekoppelt mit der kontinuierlichen spektralen Schätzung der Störung mittels Minimumstatistik nach Martin [Mar94, Mar01] und spektraler Subtraktion zur Schätzung eines Sprachreferenzsignals wird von Aichner et al. [AHBK03] vorgeschlagen.

Bei der statistischen Auswertung der durch die Kovarianzmatrizen von Sprach- und Störsignal aufgespannten Unterräume (engl. *Subspace*) im Zeitbereich oder der Matrizen der spektralen Leistungsdichten im Frequenzbereich erhält man eine gänzlich andere Klasse von Algorithmen (engl. *Subspace Approach*). Die Idee hierbei ist, eine gemeinsame Diagonalisierung der betrachteten Matrizen mit Hilfe der zugehörigen Eigenvektoren durchzuführen, um die optimalen MMSE-Filterkoeffizienten, bestehend aus den orthogonalen Matrizen dieser Eigenvektoren und der Diagonalmatrix der kombinierten Eigenwerte, zu erhalten. Die Berechnung der Eigenvektoren im Zeitbereich führt zu einer verallgemeinerten Singulärwertzerlegung (engl. *Generalized Singular Value Decomposition*, GSVD), die entweder sehr rechenintensiv pro Abtastzeitpunkt, etwas effizienter über einen Rekursionsalgorithmus nach Doclo und Moonen [DM01] oder als Teilbandimplementierung nach Spriet et al. [SMW02] erfolgen kann. Ein alternatives Vorgehen zur Komplexitätsreduzierung der Filterberechnug wird in [RM05] über eine QR-Zerlegung vorgestellt. Da diese Filterverfahren keinerlei Wissen über die Sprecherpo-

sition benötigen, ist das Sprachsignal am Ausgang des *Beamformers* auch nicht verzerrungsfrei (wie bei dem MVDR-Verfahren). In [DSWM05, CBHD06] werden daher Möglichkeiten diskutiert, um den Grad der Verzerrung zu bestimmen und konstruktiv zu verwerten. Eine Erweiterung der GSVD-Technik mit zusätzlichen adaptiven Filtern in einer GSC-Struktur ist schließlich in [DM05] beschrieben.

1.2 Wissenschaftliche Ziele dieser Arbeit

Das primäre Ziel der vorliegenden Arbeit ist die Entwicklung und Untersuchung von akustischen Strahlformungsverfahren für Sprachsignale unter Verwendung eines Optimierungskriteriums, welches auf der Maximierung des Signal-zu-Rauschleistungsverhältnisses (engl. *Signal-to-Noise Ratio*, SNR) in jedem Frequenzband basiert. Dieses Kriterium hat den Vorteil, dass keine explizite Positionsbestimmung des Sprechers notwendig ist, sondern vielmehr eine blinde Optimierung mit der impliziten Berücksichtigung der gesamten Raumimpulsantwort zwischen dem Sprecher und der Mikrophongruppe erfolgt. Diese blinde Vorgehensweise beinhaltet ebenfalls, dass die geometrische Anordnung der Mikrophone unbekannt sein kann und eine Kalibrierung der Mikrophone überflüssig ist. Bisher wurde solch ein Optimierungsansatz jedoch nur für Schmalband-Strahlformungsprobleme angewendet, bei denen die Bandbreite des Eingangssignals viel kleiner als seine Mittenfrequenz ist (z. B. in der Antennentechnik). Für die akustische Strahlformung galt das Kriterium bislang als ungeeignet, da die Maximierung des SNRs für jede betrachtete Frequenzkomponente unabhängig voneinander durchgeführt wird, und sich somit Sprachsignalverzerrungen am Ausgang des *Beamformers* einstellen. Daher werden in dieser Arbeit eigenentwickelte Verfahren aufgezeigt, welche diese Verzerrungen deutlich reduzieren können. Ein weiteres Ziel ist die Entwicklung und Anpassung von Algorithmen zur adaptiven Umsetzung des Optimierungskriteriums für verschiedene Störschallfelder. Schließlich ist noch das Ziel der Arbeit unterschiedliche Strukturen zu realisieren, um eine Optimierung hinsichtlich unterschiedlicher Stationaritätsannahmen bezüglich der Sprecherposition durchzuführen: einerseits ein *Filter-and-Sum-Beamformer* für eine schnelle Adaption und andererseits ein *Generalized Sidelobe Canceller* für eine maximale Störgeräuschunterdrückung.

Ausgangspunkt ist die Darstellung und der Vergleich grundlegender Lösungsansätze zum statistisch optimalen*Beamforming* im Frequenzbereich. Diese Ansätze sind insbesondere: *Minimum Variance Distortionless Response*, *Maximum Likelihood*, *Minimum Mean Squared Error* und die Maximierung des SNRs (Max-SNR). Dabei kommt jeweils die allgemeine Annahme einer Mehrwegeausbreitung der akustischen Signale – also die Halleigenschaft von Räumen – zum Tragen. Beim Vergleich der resultierenden Filterkoeffizienten aus den unterschiedlichen Ansätzen zeigt sich, dass sie sich gerade in einem skalaren Faktor unterscheiden. Dieser kann in Form eines einkanaligen Nachfilters realisiert werden, über diesen dann die Lösungen ineinander überführbar sind. Es werden daher drei eigenentwickelte Methoden vorgestellt, um mit Hilfe eines geeigneten Nachfilters eine approximative Realisierung eines MVDR *Beamformers* basierend auf der Maximierung des SNRs darzustellen. Somit bleiben die Vorteile des SNR-Optimierungskriteriums erhalten, wobei gleichzeitig der Nachteil der Sprachverzerrung zu einem Großteil überwunden wird.

Da bei der vorliegenden Arbeit nicht die Konzeption einer mehrkanaligen Sprachsignalverbesserung für eine konkrete Problemstellung im Vordergrund steht, werden unterschiedliche

Realisierungen für unterschiedliche Anwendungsszenarien vorgestellt. Diese hängen einerseits von dem zu erwartenden Störschallfeld und andererseits von der zu erwartenden Dynamik der Sprecherbewegung ab. Für Letztgenanntes gilt, dass bei einem sich bewegenden Sprecher eine *Filter-and-Sum-Beamformer*-Struktur mit geringen Filterlängen aufgrund der schnellen Nachführung der Filterkoeffizienten sinnvoll erscheint. Bei einer relativ statischen Anordnung hingegen ist die Struktur eines *Generalized Sidelobe Cancellers* mit größeren Filterlängen möglich, da sie zu einer höheren Rauschunterdrückung führt.

Aufgrund der Relevanz der Eigenschaften der Störung erfolgt eine Unterteilung verschiedener Störungen bzw. Störschallfelder. Die Formulierung des Optimierungskriteriums fällt je nach dem, ob gerichtete Störschallquellen vorhanden sind oder nicht, anders aus. Wird davon ausgegangen, dass keine gerichteten Störschallquellen aktiv sind, oder diese zumindest sehr wenig Leistung im Vergleich zum Sprecher emittieren, so ergibt sich das spezielle Eigenwertproblem bezüglich der Matrix der Kreuzleistungsdichten der Sprachsignale an den Mikrophonen. Der resultierende Filterkoeffizientenvektor aus der Maximierung des SNRs ist folglich gerade der dominante Eigenvektor des speziellen Eigenwertproblems. Sind starke gerichtete Störschallquellen aktiv, so ergibt sich das verallgemeinerte Eigenwertproblem bezüglich zweier Kreuzleistungsdichtematrizen: die eine beinhaltet nur die Störung und die andere enthält zusätzlich die Sprache. Daraus ergibt sich als optimaler Filterkoeffizientenvektor der dominante Eigenvektor des entsprechenden verallgemeinerten Eigenwertproblems. In dieser Arbeit werden eigenentwickelte Gradientenverfahren zur adaptiven Lösung des speziellen und des verallgemeinerten Eigenwertproblems vorgestellt. Es findet ein Vergleich zu ausgewählten Verfahren aus der Literatur statt, und die letztendlich verwendeten, modifizierten Algorithmen werden mit entsprechenden Adaptionsschemata angegeben.

Einen weiteren Schwerpunkt der Arbeit stellt die Entwicklung einer GSC-Struktur basierend auf dem verallgemeinerten Eigenwertproblem dar. Insbesondere wird eine neue *Blocking Matrix* vorgestellt, die die Vorteile besitzt, dass auch verhallte Sprachsignale in hohem Maße gedämpft werden, und dass eine Adaption auf den Sprecher hin erfolgen kann, wenn gleichzeitig ein starkes stationäres Störschallfeld vorliegt. Die Komponente des so genannten *Fixed Beamformers* wird in zwei Varianten realisiert: Zum einen mit einem DSB und zum anderen mit einem *Matched Filter*, der aus einer Modifikation des dominanten Eigenvektors hervorgeht. Der GSC mit der eigenentwickelten *Blocking Matrix* und einem idealen DSB als *Fixed Beamformer* zeigt nahezu das gleiche Leistungsverhalten wie das verwendete Referenzsystem[7]. Die Verwendung des *Matched Filters* anstatt des DSBs führt zu geringfügigen Sprachsignalverzerrungen, hat jedoch den Vorteil, keine Information über die Sprecherrichtung zu benötigen.

Weiterhin wird in der vorliegenden Arbeit gezeigt, wie mit Hilfe des adaptiv bestimmten dominanten Eigenvektors eine relativ zuverlässige Sprecherrichtungsschätzung möglich ist, obwohl starke gerichtete Störschallfelder das eigentliche Sprachsignal überlagern.

Obschon hier das primäre Ziel in der Verbesserung von Sprachsignalen liegt, bei denen zu einem gegebenen Zeitpunkt nur ein Sprecher aktiv ist, erfolgt im Anhang ein kleiner Exkurs zur blinden Quellentrennung. Dabei besteht die Problemstellung darin, zwei gleichzeitig aktive Sprecher zu trennen, also zwei Ausgangssignale zu erzeugen. Diese beinhalten dann im Idealfall jeweils nur das Signal eines Sprechers. Auch für diese Anwendung werden modifizierte

[7] Als GSC-Referenzsystem wird die Frequenzbereichsrealisierung von [HSH99] verwendet, wobei ein idealisiertes Sprachreferenzsignal zur Adaption herangezogen wird.

Adaptionsalgorithmen zur Lösung eines speziellen Eigenwertproblems verwendet.

Gliederung dieser Arbeit

Die vorliegende Arbeit lässt sich in drei Teile gliedern: Im ersten Teil (Kapitel 2 und 3) werden zuerst relevante akustische Eigenschaften geschlossener Räume erläutert, die für das Verständnis der im Folgenden untersuchten Störszenarien notwendig sind. Die Erklärungen zu einigen Begriffen der Raumakustik sind ebenfalls für die Beurteilung der Sprachsignalqualität hilfreich. Danach erfolgt eine Beschreibung möglicher Anordnungen von Mikrophongruppen und die Einführung wesentlicher Größen, welche sich aus der Richtcharakteristik ergeben. Diese sind für die frequenzabhängige objektive Messung von Leistungsmerkmalen mehrkanaliger Ansätze zur Sprachsignalverbesserung notwendig.

Der zweite Teil (Kapitel 4, 5 und 6) beschäftigt sich mit unterschiedlichen Ansätzen zum statistisch optimalen *Beamforming* und Verfahren zur iterativen Lösung des Eigenwertproblems für das SNR-Optimierungskriterium. Es werden eigenentwickelte Adaptionsvorschriften vorgestellt und experimentelle Untersuchungen zum Konvergenzverhalten präsentiert. In Kapitel 6 wird mittels neuartiger Nachfilter der Zusammenhang zwischen dem SNR-Optimierungskriterium und einem verallgemeinerten MVDR *Beamforming* hergestellt.

Der abschließende dritte Teil (Kapitel 7 und 8) behandelt die Möglichkeit einer robusten Sprecherrichtungsschätzung mittels Eigenwertzerlegung und die Realisierung eines *Generalized Sidelobe Canceller* mittels neuartiger Ansätze für die *Blocking Matrix* in Kombination mit einem *Delay-and-Sum-Beamformer* aber auch einer "blinden" Variante mit einem *Matched Filter*.

Kapitel 2

Statistische Raumakustik

Für die Beschreibung akustischer Signale, die sich am Aufnahmeort von Mikrophonen ausbilden, ist es notwendig, eine Einteilung unterschiedlicher Schallfelder durchzuführen. Dabei wird insbesondere auf die statistischen Eigenschaften der Schallfelder eingegangen, welche maßgeblich durch die raumakustischen Bedingungen bestimmt werden. In diesem Kapitel erfolgt zuerst eine Einführung in die Grundlagen der statistischen Raumakustik, wobei es im Wesentlichen um die Definition der Nachhallzeit und des Hallradius geht. Dafür wird insbesondere die Raumimpulsantwort betrachtet und deren Simulationsmöglichkeit für kleine Räume. Weiterhin erfolgt eine Analyse der Schallausbreitung in Räumen anhand der räumlichen Kohärenz sowie die Formulierung des Signalmodells, welches die Signale an den Mikrophonen des *Arrays* beschreibt. Dabei wird auf die Problematik beim Messen der räumlichen Kohärenz eingegangen. Abschließend sind einige Ergebnisse von Messungen an simulierten Schallfeldern, aber auch an Aufnahmen von Störfeldern in realen Umgebungen aufgeführt.

2.1 Schallausbreitung in Räumen

In halligen Räumen werden Schallwellen an begrenzenden Flächen und Einrichtungsgegenständen reflektiert. Daher ist es sinnvoll, eine grobe Einteilung der Schallausbreitung in Räumen vorzunehmen in die direkte Komponente, also den Direktschall von der Quelle zur Imissionsstelle, und in indirekte Komponenten aufgrund der Reflexionen. Dabei kann der indirekte Anteil noch unterteilt werden in so genannte frühe Reflexionen und den Nachhall.

Um das von einer Schallquelle erzeugte Schallfeld vollständig zu beschreiben, wäre es notwendig, für alle angeregten Frequenzen die Eigenschwingungen des Raums zu betrachten und zu überlagern. Die Schallausbreitung einzelner Frequenzkomponenten kann durch Differentialgleichungen aus der wellentheoretischen Raumakustik beschrieben werden. Streng genommen gibt es nur noch eine zweite Methode zur Analyse von Schallvorgängen, die geometrische Raumakustik. Sie bietet eine einfache Möglichkeit zur Beschreibung der Schallausbreitung im Raum in Form von geradlinigen Schallstrahlen. Da jedoch auch bei dem Modell der geometrischen Raumakustik mit fortschreitendem Beobachtungszeitraum die Komplexität drastisch steigt, können über das Schallfeld keine exakten Aussagen gemacht werden. Unter der Annahme, dass die Energiedichte des Schalls im Raum näherungsweise gleichverteilt ist, geht man zu einem dritten Modell, der so genannt statistischen Raumakustik, über. Diese beschäftigt sich nicht mit der Beschreibung aller Ausbreitungspfade der Schallstrahlen, sondern charak-

terisiert Räume durch deren Schallfeldparameter. Zwei wesentliche Größen sind hierbei zum einen die Nachhallzeit, welche die Zeitdauer beschreibt, nach der die Schallenergiedichte im Raum um einen definierten Teil gesunken ist, nachdem die Schallquelle abgeschaltet wird. Zum anderen ist dies der Hallradius, der die Entfernung angibt, bei der die direkte gleich der reflektierten Schallenergie ist.

Betrachtet man Schallwellen mit einer gewissen Anfangsenergie E_0, welche sich im Raum ausbreiten, so wird die Energie nach jeder Reflexion abnehmen und der Zeitverlauf der Energiedichte nimmt die Exponentialform

$$E(t) = E_0 e^{-\frac{t}{\tau}} \tag{2.1}$$

an. Die zeitliche Dämpfungseigenschaft des Raums τ wird üblicherweise durch die Nachhallzeit ausgedrückt, die wiederum definiert ist als die Zeitdauer, in der die Schallenergie auf ein Millionstel gesunken ist bzw. der Schalldruckpegel um 60 dB vom Anfangswert abfällt [Sab22]. Daher wird die Nachhallzeit auch häufig mit T_{60} benannt. Sie ist die bekannteste und wohl wichtigste raumakustische Kenngröße. Für die Dämpfungskonstante ergibt sich somit

$$\tau = -\frac{T_{60}}{\ln(10^{-6})}. \tag{2.2}$$

Für den stationären Zustand, wenn dem Raum vom Volumen V die konstante Schallleistung P zugeführt wird, lässt sich diese angeben zu

$$P = \frac{\ln(10^6) V p^2}{(1-\bar{\alpha}_A) T_{60} c^2}. \tag{2.3}$$

Hierbei gibt p den Schalldruck und c die Wellengeschwindigkeit an. Der absorbierte Schallteil durch die Raumoberflächen ist mit dem mittleren Absorptionsgrad[1] $\bar{\alpha}_A$ bezeichnet. In der Praxis ist häufig ein einfacher geometrischer Zusammenhang zwischen der Nachhallzeit und dem Absorptionsgrad in analytischer Form von großer Bedeutung. Dafür kann eine mittlere freie Weglänge $\bar{l} = 4V/A$ des Schallstrahls im Raum mit dem Volumen V und der Wandfläche A angesetzt werden [CM78]. So ergibt sich eine mittlere Stoßzahl

$$\bar{n} = c/\bar{l} = \frac{Ac}{4V}, \tag{2.4}$$

welche die Anzahl der Reflexionen des Schalls pro Zeit angibt. Mit dieser lässt sich folgender zeitlicher Abfall der Schallenergiedichte angeben

$$E(t) = E_0 (1-\bar{\alpha}_A)^{\frac{Ac}{4V}t} = E_0 e^{\frac{Ac\ln(1-\bar{\alpha}_A)}{4V}t}. \tag{2.5}$$

Hierbei ist die Dämpfung im Luftvolumen während der Wellenausbreitung unberücksichtigt geblieben. Ein Vergleich von Gl. (2.1) und Gl. (2.2) mit Gl. (2.5) liefert schließlich den gewünschten Zusammenhang[2]

$$T_{60} = \frac{4\ln(10^{-6})V}{Ac\ln(1-\bar{\alpha}_A)}. \tag{2.6}$$

[1] Mit $\bar{\alpha}_A$ ist der mittlere Absorptionsgrad in einem Raum und mit α_A der Absorptionsgrad einer homogenen Fläche bezeichnet. Komplementär zu $\alpha_A = 1 - \rho_R$ ist der Reflexionsgrad ρ_R, mit $\alpha_A, \rho_R \in [0, \ldots, 1]$. Legt man anstelle von Energien Amplituden zugrunde, spricht man von Faktoren: Reflexionsfaktor und Absorptionsfaktor. Diese können dann auch negative Werte annehmen, wodurch Phasendrehungen berücksichtigt werden.

[2] Häufig findet man in der Literatur die Nachhallformel nach Sabine, in der die Vereinfachung $\ln(1-\bar{\alpha}_A) \approx -\bar{\alpha}_A$ im Falle kleiner und mittlerer Absorptionsgrade eingesetzt wird. Diese ist jedoch nur zur Beschreibung großer Räume zulässig.

Diese einfache Näherung ist noch im folgenden Abschnitt von Bedeutung, wenn es um die Simulation der Schallausbreitung innerhalb eines definierten Raumes geht, wobei eine gewünschte Nachhallzeit vorgegeben werden soll. Es ist offensichtlich, dass für ein genaues Verhältnis zwischen der Raumbeschaffenheit und der Nachhallzeit sich die Gesamtfläche aus Teilflächen mit unterschiedlichen Reflexionskoeffizienten ergibt [Eyr30]. Noch problematischer ist allerdings, dass die Nachhallzeit in der Realität frequenzabhängig ist. Dieser Umstand wird dadurch hervorgerufen, dass der Absorptionsgrad α_A eines Materials nicht für jede Schallfrequenz derselbe ist. In aller Regel sinkt dieser mit abfallender Frequenz. Während hohe und teilweise auch mittlere Tonlagen noch recht gut von Materialien mit hohem α_A gedämpft werden, hat das gleiche Material im Bereich tiefer Frequenzen praktisch keine Auswirkungen mehr auf den Schall. Durch die Frequenzabhängigkeit der Nachhallzeit werden manche Frequenzanteile eines Geräusches länger zum Ausklingen benötigen, als andere Teile. Diese Effekte werden jedoch häufig in der Raumakustik nicht berücksichtigt.

Je nach raumakustischem Zweck können unterschiedliche optimale Nachhallzeiten angegeben werden. Bei Aufnahme- und Regieräumen z. B. sind sehr niedrige Nachhallzeiten von $T_{60} < 0{,}2\,\mathrm{s}$ notwendig. Für Büroräume ist ebenfalls eine geringe bis mittlere T_{60}-Zeit von $0{,}3\,\mathrm{s}$ bis $0{,}5\,\mathrm{s}$ üblich und für Vortragssäle bereits höhere zwischen $0{,}6\,\mathrm{s}$ und $0{,}8\,\mathrm{s}$. Bei Räumen für Musikdarbietung hängt die optimale Nachhallzeit von der Art der Darbietung ab. Sie kann Werte zwischen $1\,\mathrm{s}$ und $3\,\mathrm{s}$ annehmen.

Aufgrund der vielfachen Reflexionsmöglichkeiten in verhallten Räumen trifft der Nachhall an einem Raumpunkt mit zunehmender Laufzeit aus allen Richtungen mit ähnlicher Intensität ein. Allerdings weist erst der späte Nachhall im Idealfall eine konstante Schallenergiedichte im Raum auf (isotrop) [CM78]. Solch ein Schallfeld wird daher auch als diffuses Schallfeld bezeichnet und hat in der Praxis eine besondere Bedeutung. In unmittelbarer Umgebung einer Schallquelle herrschen ähnliche Bedingungen wie im Freien, die Raumrückwirkungen machen sich erst mit zunehmendem Abstand bemerkbar.

Das Direktschallfeld kann näherungsweise durch Kugelwellenausbreitung beschrieben werden, d. h. die Energiedichte verhält sich reziproportional zu dem Quadrat der Entfernung r vom Sender gemäß:

$$E_D = \frac{P}{4\pi r^2 c}. \tag{2.7}$$

Für das stationäre Schallfeld gilt hingegen mit der Beziehung $E = p^2/((1-\bar{\alpha}_A)c^2)$ und Gl. (2.3)

$$E_{St} = \frac{PT_{60}}{\ln(10^6)V}. \tag{2.8}$$

Der Hallradius r_H ist nun jener Abstand, bei dem die stationäre Energiedichte gleich der des Direktschallfeldes ist

$$r_H = \sqrt{\frac{\ln(10^6)V}{4\pi c T_{60}}}. \tag{2.9}$$

Die sich hieraus ergebenden Hallradien sind allerdings erstaunlich gering. So würde nur für in der Nähe zum Sender aufgestellte Mikrophone die Energiedichte des Direktschalls die statistische Energiedichte überwiegen. Nun haben jedoch nur wenige Schallquellen eine allseitig gleichmäßige Energieabstrahlung. Im Allgemeinen ist mit einer ausgeprägten Richtwirkung zu rechnen, welche durch einen Korrekturterm unter der Wurzel in Gl. (2.9) berücksichtigt wird.

Dieser so genannte Bündelungsgrad gibt das Verhältnis der Schallintensität in Hauptstrahlrichtung zu deren Mittelwert über alle Richtungen an und kann Werte bis zu 100 annehmen. Bei einem Sprecher als Schallquelle ergibt sich z. B. ein Bündelungsgrad von ungefähr 2 [Mar95]. Praktisch bedeutet dies, dass ein Redner in seiner direkten Nähe gut zu verstehen und der Nachhall kaum wahrnehmbar ist. Weiter entfernt wird diese Stimme immer mehr im Nachhall untergehen und die Verständlichkeit nimmt deutlich ab. Aus nur einem Mikrophonsignal jenseits des Hallradius ist die Richtung des Direktsignals nicht mehr eindeutig bestimmbar, wenn man es um seine Achse dreht. Anders verhält sich dagegen das menschliche binaurale Hören, das es uns ermöglicht noch weit außerhalb des Hallradius die Richtung der Schallquelle zu bestimmen.

In Bild 2.1 ist der Übergang von Direktschall zu diffusem Schall anhand des relativen Schalldruckpegels L_{rel} dargestellt [Dic97]. Zu erkennen ist hierbei, dass der Pegel des Direktschalls um 6 dB je Verdoppelung des Abstandes zwischen Schallquelle und Empfänger abnimmt und der Gesamtschallpegel mit steigender Entfernung auf den Diffusschallpegel sinkt.

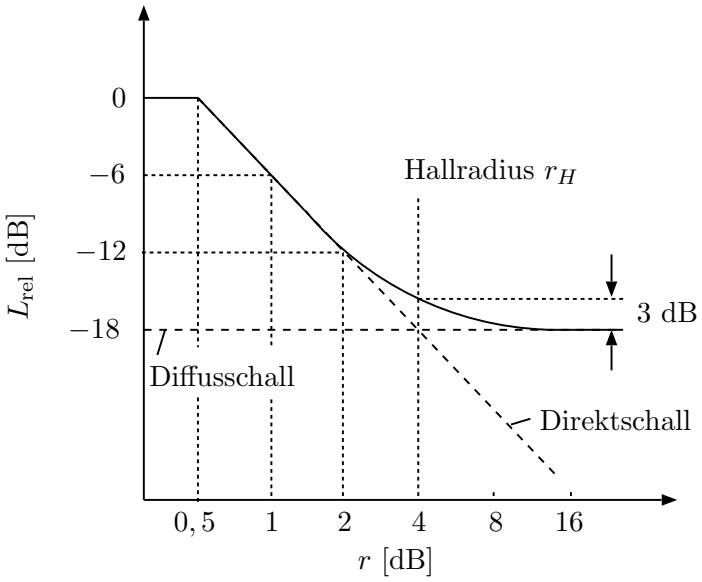

Bild 2.1: Überlagerung von Direkt- und Diffusschall.

2.2 Raumimpulsantworten

Die bisher beschriebenen Schalleigenschaften sollen nun anhand der Raumimpulsantwort betrachtet werden. Theoretisch bewirkt ein einzelner Impuls der Schallquelle am Ort des Empfängers aufgrund der Reflexionen eine ganze Folge von Impulsen, deren Dichte mit der Zeit zunimmt und deren Amplitude immer geringer wird. Jede Raumimpulsantwort ist spezifisch für den Raum und für die verwendete Sender- Empfängeranordnung. Bild 2.2 zeigt schematisch eine solche Impulsantwort, wobei eine Aufteilung in charakteristische Teilbereiche vorgenommen wurde. Der zeitlich erste Impuls wird dem Direktschall zugeordnet, da er den kürzesten Ausbreitungsweg von der Quelle zum Empfänger nimmt. Nach Gl. (2.7) ist dessen Amplitude dabei umso kleiner, je weiter die Schallquelle vom Empfänger entfernt ist. Dem Direktschall folgen die frühen Echos, welche auf Schallanteile mit nur wenigen Wandreflexionen

zurückzuführen sind. Aufgrund der geringen Verzögerung gegenüber dem Direktschall können diese Reflexionen und der direkte Schall nicht vom Ohr unterschieden werden, weshalb sie die Verständlichkeit von Sprache (Def. siehe unten) und die Transparenz von Musik erhöhen. Der sich anschließende frühe Nachhall geht bereits auf vermehrte Wandreflexionen zurück, ist jedoch noch richtungsabhängig. Daher trägt er zu einem räumlichen Klangeindruck bei. Im Bereich des späten Nachhalls ist keine Unterscheidung einzelner Echos mehr möglich, da eine gleichmäßige Verteilung der Schallleistung über den gesamten Raum vorliegt. Erst in diesem Bereich klingt die Intensität nach dem Prinzip der statistischen Raumakustik exponentiell ab [CM78], vgl. Gl. (2.1).

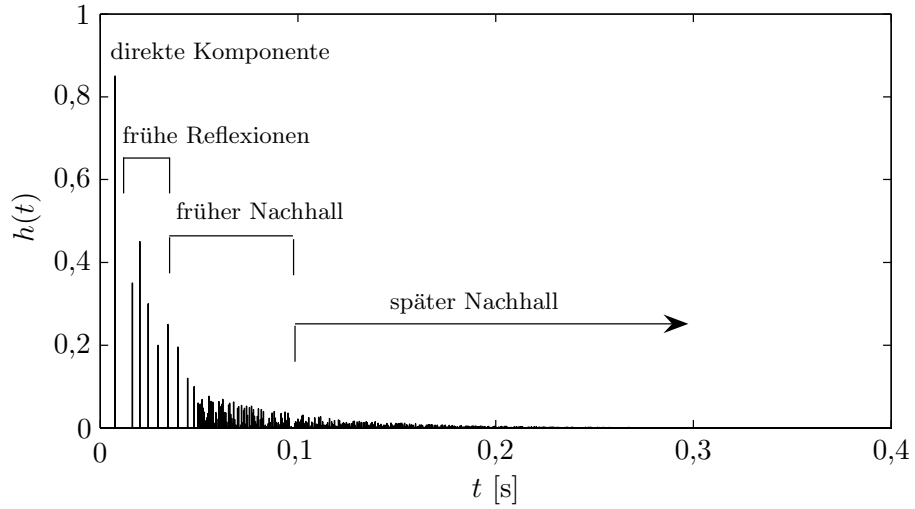

Bild 2.2: Schematische Darstellung einer Raumimpulsantwort $h(t)$.

Es existieren eine Vielzahl von Termini, mit Hilfe derer eine Einschätzung von Sprach- bzw. Musikwiedergabe erfolgt. Dabei ist es möglich, dass ein und derselbe Begriff je nach Literatur eine unterschiedliche Bedeutung hat. In z. B. [Ber96] findet sich eine umfangreiche Begriffsbestimmung[3] zu dieser Thematik. Studien ab den 1960er Jahren haben schließlich zu den heute gängigen Gütemaßen geführt, die es ermöglichen, numerische Aussagen über die akustische Raumqualität[4] zu geben. An dieser Stelle soll lediglich auf die allgemeine Verständlichkeit von Sprache eingegangen werden, welche durch die Anfangsnachhallzeit und das Deutlichkeitsmaß charakterisiert sind. Basierend auf dem Vergleich von frühen und späten Anteilen der Impulsantwort wurde bereits in [Thi53] folgendes Kriterium für den relativen Anteil an nützlichem Schall vorgeschlagen:

$$\vartheta(t_g) = \frac{\int_0^{t_g} h^2(t)dt}{\int_0^{\infty} h^2(t)dt}, \tag{2.10}$$

wobei $\vartheta(t_g = 50\,\mathrm{ms})$ Deutlichkeit genannt wurde. Aus Gl. (2.10) hat sich das heute übliche

[3]Die detaillierte Beschreibung des akustischen Eindrucks bei Sprach- und Musikwiedergabe in z. B. Konzerträumen ist für Raumakustiker aber auch für Toningenieure von wichtiger Bedeutung. Dafür existiert ein umfangreiches Vokabular wie z. B. Abstimmung, Brillanz, Flimmern, Intimität und viele andere.

[4]Für die Beurteilung von Räumen für die musikalische Darbietung werden mehrere Gütemaße und deren Kombination verwendet, die jeweils auf der Energie der Raumimpulsantwort für verschiedene Zeitintervalle basieren. So ergeben sich Größen wie z. B. Seitenschallgrad, Bass-Verhältnis oder Silbenverständlichkeit.

Deutlichkeitsmaß C_{50} für Sprache und das Klarheitsmaß C_{80} für Musik ergeben:

$$C_{50} = 10\log\frac{\int_0^{50\ ms} h^2(t)dt}{\int_{50\ ms}^{\infty} h^2(t)dt}, \qquad C_{80} = 10\log\frac{\int_0^{80\ ms} h^2(t)dt}{\int_{80\ ms}^{\infty} h^2(t)dt}. \tag{2.11}$$

Die Wahl von $t_g = 50$ ms ist durch den psychoakustischen Effekt der Trägheit des Ohres begründet, der besagt, dass Impulse, die weiter als diese Zeit auseinander liegen, erst einzeln erkennbar sind (vgl. Einteilung der Raumimpulsantwort in Bild 2.2). Eine weitere psychoakustische Auswirkung ist die so genannte Verdeckung. Dabei werden Töne frequenzselektiv verdeckt, welche unterhalb eines gewissen Schallpegels relativ zu einem zusätzlich vorhandenen energiereicheren Ton auftreten. D. h., dass für das menschliche Empfinden des Nachhalls vor allem der Anfangsteil des Abklingvorgangs deutlicher wahrgenommen wird als der spätere Bereich der Nachhallzeit, da diese normalerweise durch nachfolgende Töne überdeckt wird. In [Jor74] wurde daher die Anfangsnachhallzeit T_A definiert (engl. *Early Decay Time*, EDT). Diese gibt die Zeit an, in welcher die Schallintensität um 10 dB abnimmt:

$$-10\ \text{dB} \stackrel{!}{=} 10\log\frac{\int_0^{T_A} h^2(t)dt}{\int_0^{\infty} h^2(t)dt}\ \text{dB}. \tag{2.12}$$

Die Größen für die Anfangsnachhallzeit sowie für die Nachhallzeit können anschaulich aus der Darstellung der Rückwärtsintegration der quadrierten Raumimpulsantwort ersehen werden (Schröder-Rückwärtsintegration) [Sch65]. Häufig wird eine so ermittelte Energieabfallkurve (engl. *Energy Decay Curve*, EDC) in der normierten Form angegeben:

$$E_A(t) = 10\log\frac{\int_t^{\infty} h^2(t)dt}{\int_0^{\infty} h^2(t)dt}\ \text{dB}. \tag{2.13}$$

In Bild 2.3 ist eine Energieabfallkurve beispielhaft für die Impulsantwort aus Bild 2.2 mit einer Anfangsnachhallzeit $T_A = 46$ ms, einer Nachhallzeit $T_{60} = 348$ ms und einem Deutlichkeitsmaß $C_{50} = 9{,}9$ dB dargestellt.

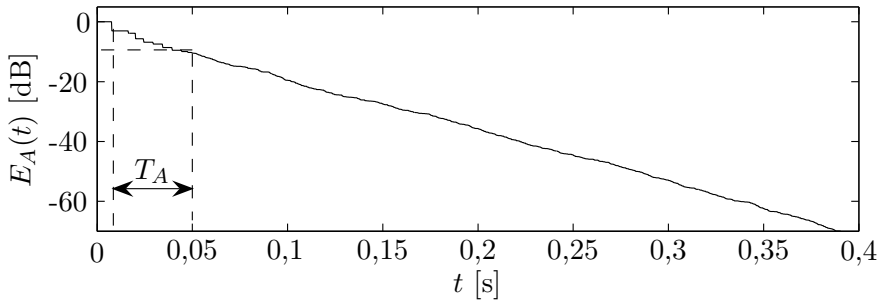

Bild 2.3: Energieabfallkurve durch Schröder-Rückwärtsintegration mit einer Anfangsnachhallzeit von $T_A = 46$ ms.

Simulation der Schallausbreitung

Zur Bestimmung von Raumimpulsantworten in simulierten Räumen können grundsätzlich zwei Arten von Verfahren verwendet werden. Zum einen sind dies wellentheoretische Ansätze und zum anderen geometrische Verfahren. Als geeignetes Werkzeug für die wellentheoretische Behandlung und modale Analyse von akustischen Problemen hat sich in den letzten Jahren

die Methode der Finiten Elemente (engl. *Finite Element Method*, FEM) bewährt [Bar03]. FE-Verfahren kommen zum Einsatz, wenn bei der Berechnung Phaseneffekte von Schallfeldern eine Rolle spielen, d. h. wenn Moden, Beugung oder Streuung zu berücksichtigende Effekte sind.

Bei den geometrischen Verfahren zur analytischen Bestimmung von Raumimpulsantworten stellen einerseits das Schallteilchenverfahren und andererseits die Spiegelquellenmethode die wichtigsten Methoden dar. Weiterhin werden in der Raum- und Bauakustik zur möglichst genauen Simulation der Schallausbreitung Kombinationen beider Verfahren in Hybrid-Methoden eingesetzt, um die jeweiligen Vorteile beider Verfahren zu nutzen [RC03], [Bar03]: die Spiegelquellenmethode ist ein schnelles und exaktes Verfahren zur Berechnung des ersten Teils der Raumimpulsantwort und eignet sich insbesondere für nichtgekrümmte Oberflächen. Bei dem Schallteilchenverfahren liegt der Vorteil in der effizienteren Berechnung des späteren Verlaufs der Raumimpulsantwort, sowie der Analyse gekrümmter Oberflächen im simulierten Raum. In beiden Fällen wird üblicherweise von einer punktförmigen und radial abstrahlenden Schallquelle ausgegangen.

Bei der Schallteilchenmethode (auch *Ray-Tracing*-Verfahren genannt) werden in zufällig ausgewählte Richtungen Teilchen ausgesendet, die mit einer Anfangsenergie und einem Zeitstempel versehen sind. Sie werden an den Wänden reflektiert und verlieren je nach Oberflächeneigenschaften einen Teil ihrer Energie. Von jedem Teilchen, das am Empfänger eintrifft wird dann die verbleibende Energie und die Ausbreitungszeitdauer in die Impulsantwort "eingetragen".

Das Spiegelquellenmodell bietet eine sehr effiziente Methode zur Simulation der Ausbreitung eines Schallfeldes in Räumen einfacher Geometrie und geringer Nachhallzeit, welche insbesondere häufig im Bereich der Sprachsignalverarbeitung verwendet wird [AB79]. Dabei treffen die von der Schallquelle (Sprecher) emittierten Kugelschallwellen am Empfänger einerseits auf direktem Wege, andererseits über Reflexionen durch die Wände an. Der beim Empfänger erzeugte Schalldruck hängt nur von der Entfernung zum Sender ab, nicht aber vom Einfallswinkel. Daher kann von jeder reflektierten Welle angenommen werden, dass sie einer virtuellen Kugelschallquelle, deren Entfernung vom Empfänger der Lauflänge des Schalls entspricht, entsprungen ist, welche durch Spiegelung der Schallquelle an den Raumbegrenzungen entstanden ist. Die Ordnung einer Spiegelquelle gibt an, wie oft der durch sie repräsentierte Schallstrahl reflektiert wird, bevor er den Empfänger erreicht, siehe Bild 2.4 (a) für ein Beispiel erster und zweiter Ordnung. Dabei hängt die Amplitude des reflektierten Signals von der Wandabsorption ab, welche sich z. B. bei der vereinfachten Annahme gleicher Beschaffenheit der Oberflächen und Vorgabe einer bestimmten Nachhallzeit aus Gl. (2.6) berechnen lässt. Die Gesamtanordnung wird also derart bestimmt, dass die Position der Quelle an den Wänden des Raumes dreidimensional gespiegelt und alle entstandenen Spiegelquellen als neue Schallquellen interpretiert werden, siehe Bild 2.4 (b).

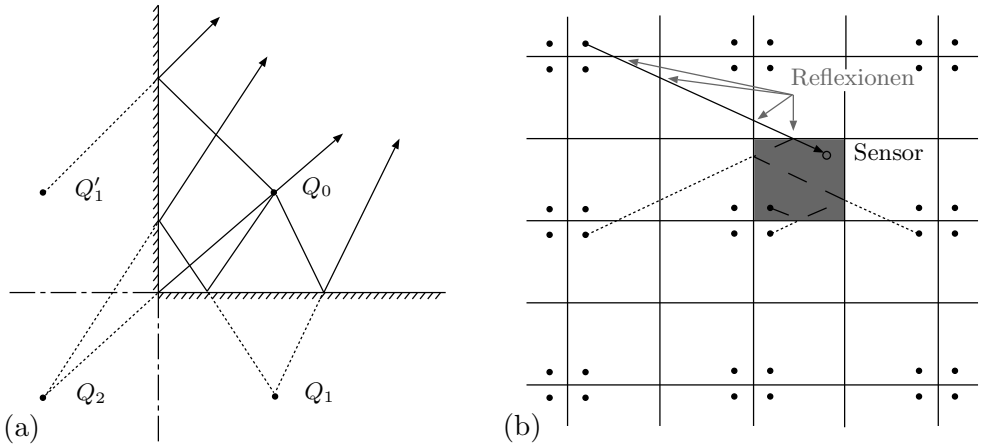

Bild 2.4: In (a) beispielhafter Verlauf der Originalwege sowie der virtuellen Wege der Schallausbreitung erster und zweiter Ordnung. In (b) Erweiterung des Raums nach der Spiegelquellenmethode zur Demonstration Reflexionen vierter Ordnung.

Im Verlauf dieser Arbeit wurde eine Vielzahl von akustischen Signalen analysiert, welche aufgrund von mit simulierten Raumimpulsantworten gefalteten Sprachsignalen hervorgegangen sind. Da die hierbei betrachteten Räume als relativ klein angenommen werden und von einfacher geometrischer Beschaffenheit sind – also keine konkrete Anordnung exakt nachgebildet werden soll – wurde die Spiegelquellenmethode zur Erzeugung von Raumimpulsantworten benutzt. Dabei wird im Allgemeinen von einem quaderförmigen, leeren Raum ausgegangen, in dem sich eine punktförmige Schallquelle und mehrere punktförmige Schallempfänger befinden.

2.3 Mehrkanaliges Signalmodell

Im letzten Abschnitt wurde auf die Eigenschaften des Direktschallfelds, früher Reflexionen und des Nachhalls eingegangen. Mit Hilfe einer gegebenen Raumimpulsantwort können so aus der Energieabfallkurve akustische Raumeigenschaften abgelesen werden. Nun soll der Fall betrachtet werden, dass Aussagen über ein gegebenes Schallfeld gemacht werden sollen ohne die Raumimpulsantwort zu kennen, d. h., ob der frühe Anteil der Impulsantwort dominiert oder der späte Nachhall. Dazu soll zunächst in diesem Abschnitt das dafür notwendige mehrkanalige Signalmodell zur Aufnahme der zu analysierenden akustischen Situation erfolgen. In Bild 2.5 ist hierfür eine allgemeine Anordnung von einem Sprecher, der das Signal $s_c(t)$ erzeugt, einer Störgeräuschquelle $n_c(t)$ und M akustischen Sensoren dargestellt. Der Index "c" soll andeuten, dass die durch das Nutzsignal und die Störquelle hervorgerufenen akustischen Signale an den Sensoren korreliert[5] sind, im Gegensatz zu den Signalen $n_{u,1}(n), \ldots, n_{u,M}(n)$ welche Störsignale darstellen, die untereinander als unkorreliert angenommen werden sollen. Hier soll weiterhin gelten, dass die Sensoren die Mikrophone inklusive Verstärkung und Abtastung zu den Zeiten nT modellieren, so dass anschließend zeitdiskrete Signale[6] mit der Abtastrate $1/T$

[5]Grundsätzlich wird auf die räumliche Korrelation von Schallfeldern noch in Abschnitt 2.4 eingegangen. Hier soll jedoch noch angemerkt sein, dass sowohl $s_c(t)$ als auch $n_c(t)$ frequenzabhängige räumlich unkorrelierte Komponenten im oberen Frequenzbereich bedingt durch die Eigenschaften diffuser Schallfelder (siehe Abschnitt 2.4) an den Orten der Mikrophone verursachen.

[6]Genau genommen geht bei einem zeitkontinuierlichen Signal der Parameter t nach einer Abtastung mit der Periode T in den zeitdiskreten Parameter nT über. In dieser Arbeit sollen die zeitdiskreten Signale jedoch

und dem Zeitindex n vorliegen. Die unkorrelierten Störungen $n_{u,i}(n)$ mit $i = 1, \ldots, M$ fassen das Rauschen durch die Mikrophone und die Verstärkung zusammen. Die korrelierten Signale am Aufnahmeort der Mikrophone ergeben sich aus dem Faltungsprodukt der von den Punktschallquellen abgestrahlten Signale und der zwischen Quelle und Mikrophon bestehenden Raumimpulsantwort; einerseits für das Sprachsignal mit den zeitdiskreten Impulsantworten $h_i(n)$ und andererseits für die Störquelle mit den zeitdiskreten Impulsantworten $a_i(n)$.

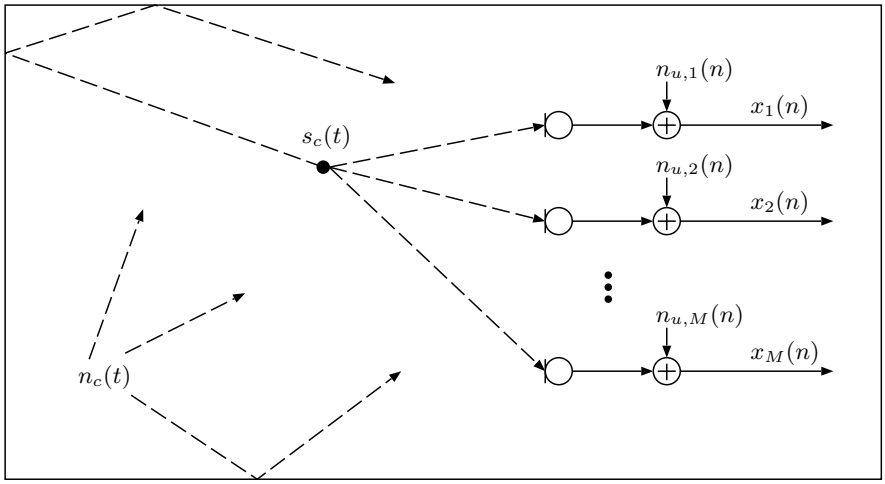

Bild 2.5: Modell zur mehrkanaligen Aufnahme von akustischen Signalen.

Der funktionale Zusammenhang zwischen den zu verarbeitenden zeitdiskreten Signalen $x_i(n)$ und den von den Quellen in dem Raum abgestrahlten Signalen kann schließlich geschrieben werden als

$$x_i(n) = s_c(n) * h_i(n) + a_i(n) * n_c(n) + n_{u,i}(n) \tag{2.14}$$

$$= s_c(n) * h_i(n) + n_i(n), \tag{2.15}$$

wobei alle Störungen im i-ten Signalpfad zu $n_i(n)$ zusammengefasst sind und $*$ den Faltungsoperator bezeichnet. Hierbei ist leicht zu ersehen, dass bei einer Erweiterung des Modells um zusätzliche Störquellen im Raum, sich diese in äquivalenter Schreibweise zu $n_c(n)$ in Gl. (2.14) additiv dem Gesamtsignal überlagern und sich schlussendlich ebenfalls alle Rauschterme wie in Gl. (2.15) zusammenfassen lassen. An dieser Stelle seien noch zwei häufig gemachte Annahmen erwähnt. Zum einen ist dies die bereits erwähnte Modellierung der Schallquellen als Punktquellen, obwohl sie genau genommen räumlich ausgedehnte Quellen sind. Zum anderen sind die Raumimpulsantworten als zeitinvariant vorausgesetzt, was zumindest bei einem Sprecher als Quelle nicht grundsätzlich eingehalten werden kann, da schon durch leichte Kopfbewegungen die Impulsantwort zur zeitlich veränderlichen Funktion wird. Vielfach wird zusätzlich angenommen, dass die Sprecherposition auf Orte in der Nähe der Mikrophone eingegrenzt werden kann, wodurch der Sprecher sich im Hallradius oder dessen Nähe befindet und somit der Direktschall dominiert.

Zeitdiskrete Fourier-Transformation Da im weiteren Verlauf die Signalbeschreibung in der Regel im Frequenzbereich erfolgt, soll hier zuerst die Darstellung der Signale im frequenzkontinuierlichen Spektrum eingeführt werden. Die Eingangssignale $x_i(n)$ erfahren dabei eine

der Einfachheit halber dimensionslos verwendet werden, z. B. $s_c(n)$ anstatt $s_c(nT)$.

zeitdiskrete Fourier-Transformation (engl. *Discrete Time Fourier Transform*, DTFT), so dass sich die entsprechenden Signale $X_i(\Omega)$ mit der normierten Kreisfrequenz Ω ergeben. Hierbei gilt der Zusammenhang $\Omega = 2\pi f / f_{Ab}$ für die betrachtete Frequenz f mit der Abtastfrequenz $f_{Ab} = 1/T$. Die gleiche Darstellung gilt natürlich ebenso für die DTFT des Rauschterms $N_i(\Omega)$ und des Sprachsignals $S_c(\Omega)$, sowie die DTFTs der entsprechenden Impulsantworten $H_i(\Omega)$ und $A_i(\Omega)$.

Diskrete Frequenzauflösung Für die Verarbeitung von Signalen im Frequenzbereich ist die Betrachtung diskreter Spektralkomponenten mit Hilfe der diskreten Fourier-Transformation (engl. *Discrete Fourier Transform*, DFT) unumgänglich. Dabei wird das kontinuierliche Spektrum an den Frequenzen $f_k = f_{Ab}/L \cdot k$ bzw. $\Omega_k = 2\pi/L \cdot k$ betrachtet, wobei $k = 0, \ldots, L-1$ gilt und L die Länge der DFT angibt. Die Eingangssignale $x_i(n)$ müssen dafür jeweils zu Segmenten der Länge L zusammengefasst und transformiert werden. Als Segmentindex soll hier der Zähler m dienen, so dass sich für den m-ten Block und das i-te Mikrophonsignal z. B. das diskrete Spektrum $X_{i,m}(\Omega_k)$ mit dem Frequenzindex $k = 0, \ldots, L-1$ ergibt. Die DFT wird auf digitalen Rechnern üblicherweise mit Hilfe der so genannten schnellen Fourier-Transformation (engl. *Fast Fourier Transform*, FFT) umgesetzt.

2.4 Räumliche Kohärenz akustischer Schallfelder

Von an unterschiedlichen Orten in einem Raum aufgenommenen akustischen Signalen kann man eine Kreuzkorrelation berechnen. Diese ist abhängig vom Abstand der Mikrophone und der zeitlichen Verschiebung der Signale zueinander. Daher lassen sich Schallfelder durch eine Raum-Zeit-Kreuzkorrelationsfunktion beschreiben. Die wohl bekannteste Größe bezüglich der räumlichen Korrelation von Schallfeldern ist die so genannte komplexe Kohärenzfunktion. Sie ist definiert als das Verhältnis des Kreuzleistungsdichtespektrums $\phi_{X_i X_l}(\Omega)$ zur Wurzel aus dem Produkt der Autoleistungsdichtespektren $\phi_{X_i X_i}(\Omega)$ und $\phi_{X_l X_l}(\Omega)$ für die beiden Signale $x_i(n)$ und $x_l(n)$ [BP66], [Gar92]:

$$\gamma_{X_i X_l}(\Omega) = \frac{\phi_{X_i X_l}(\Omega)}{\sqrt{\phi_{X_i X_i}(\Omega)\phi_{X_l X_l}(\Omega)}}. \tag{2.16}$$

Häufig wird aber auch das Betragsquadrat der Kohärenzfunktion (engl. *Magnitude Squared Coherence*, MSC) als Kohärenz bezeichnet

$$\Gamma_{X_i X_l}(\Omega) = \frac{|\phi_{X_i X_l}(\Omega)|^2}{\phi_{X_i X_i}(\Omega)\phi_{X_l X_l}(\Omega)}. \tag{2.17}$$

Die Kohärenz nach Gl. (2.17) nimmt in der Regel[7] nur die Werte zwischen Null und Eins an:

$$0 \leq \Gamma_{X_i X_l}(\Omega) \leq 1. \tag{2.18}$$

Für den Fall von unkorrelierten Signalen wie z. B. $n_{u,i}(n)$ und $n_{u,l}(n)$ in Gl. (2.14) wird die Kohärenz gerade zu Null. Ansonsten markiert das diffuse Schallfeld die untere Grenze der Kohärenz. Die obere Grenze ist durch den Direktschall gegeben[8] und für den theoretischen

[7] Streng genommen gilt Gl. (2.18) für zwei zeitdiskrete, verbundstationäre und mittelwertfreie stochastische Prozesse nur unter gewissen Voraussetzungen, vgl. [BP80].

[8] Die obere Grenze des Betrags der Kohärenzfunktion wird z. B. auch für den Fall korrelierter Quellen erreicht.

Fall von gegeneinander rein verzögerter Signale wird sie zu Eins, wobei dann die komplexe Kohärenzfunktion gegeben ist durch (siehe Anhang B)

$$\gamma_{X_i X_l}(\Omega) = \cos\left(\Omega/T \cdot \sin\theta \cdot \frac{d_{il}}{c}\right) + j \cdot \sin\left(\Omega/T \cdot \sin\theta \cdot \frac{d_{il}}{c}\right), \tag{2.19}$$

wobei in Gl. (2.19) mit j die imaginäre Einheit bezeichnet ist. Die Schallquelle soll sich dabei in ausreichender Entfernung[9] zu den Mikrophonen befinden, so dass die sich dann ergebende ebene Schallwelle mit dem Einfallswinkel θ auf die Sensoren trifft, welche im Abstand d_{il} zueinander angeordnet sind.

Für das diffuse Schallfeld lässt sich die Kohärenz unter der Annahme von Mikrophonen mit Kugelcharakteristik als Funktion des Mikrophonabstands geschlossen berechnen [CWB+55], [Kut00] zu

$$\Gamma_{X_i X_l}(\Omega) = \frac{\sin^2\left(\Omega/T\frac{d_{il}}{c}\right)}{\left(\Omega/T\frac{d_{il}}{c}\right)^2} = \mathrm{si}^2\left(\Omega/T\frac{d_{il}}{c}\right). \tag{2.20}$$

Bild 2.6 zeigt diesen Verlauf über der kontinuierlichen Frequenz für vier Mikrophonabstände d_{il}. Es ist zu erkennen, dass die Kohärenz bei tiefen Frequenzen bis zur ersten Nullstelle der si^2-Funktion hoch ist und mit zunehmender Frequenz und zunehmendem Mikrophonabstand schnell abnimmt.

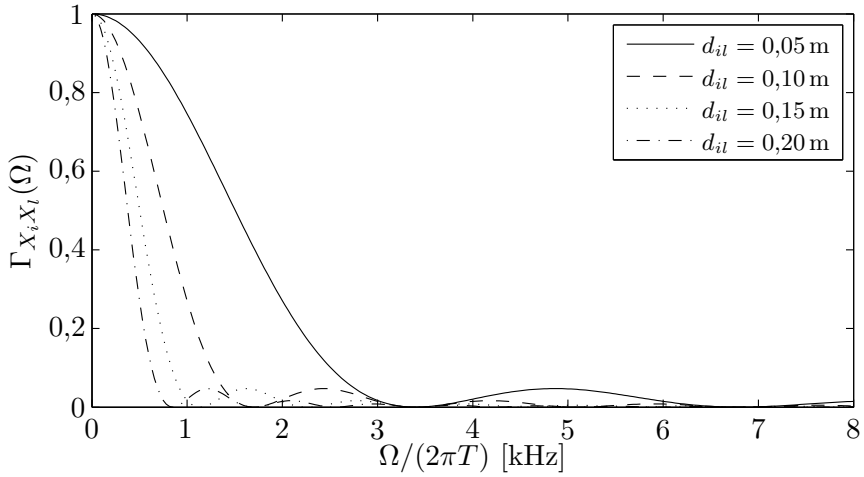

Bild 2.6: Kohärenzverlauf eines idealen diffusen Schallfeldes für unterschiedliche Sensorabstände d_{il}.

Eine Möglichkeit, ein diffuses Schallfeld zu erzeugen, ist die Anordnung unendlich vieler voneinander unabhängiger Punktschallquellen im Raum bei beliebiger Nachhallzeit. Die dabei abgestrahlten Signale weisen zueinander keine zeitlichen Korrelationen auf und die an zwei Raumpunkten aufgenommenen Signale zeichnen sich lediglich durch stochastische Phasenbeziehungen aus [DDP88]. In Bild 2.7 ist beispielhaft die gemessene Kohärenz für unterschiedliche Sensorabstände bei der kugelförmigen Anordnung einer großen Anzahl an punktförmigen unabhängigen weißen Rauschquellen um die Messpunkte herum dargestellt (siehe Anhang B).

[9] Im Falle von großen Entfernungen zwischen der Schallquelle und den Mikrophonen fällt die kugelförmig emittierte Schallwelle näherungsweise planar auf die Sensoren. Diese Näherung wird als so genannte Fernfeldnäherung bezeichnet.

Dabei erfolgte die Messung[10] der Leistungsdichtespektren und der Kohärenz abschnittweise, was im Folgenden noch genauer betrachtet werden soll.

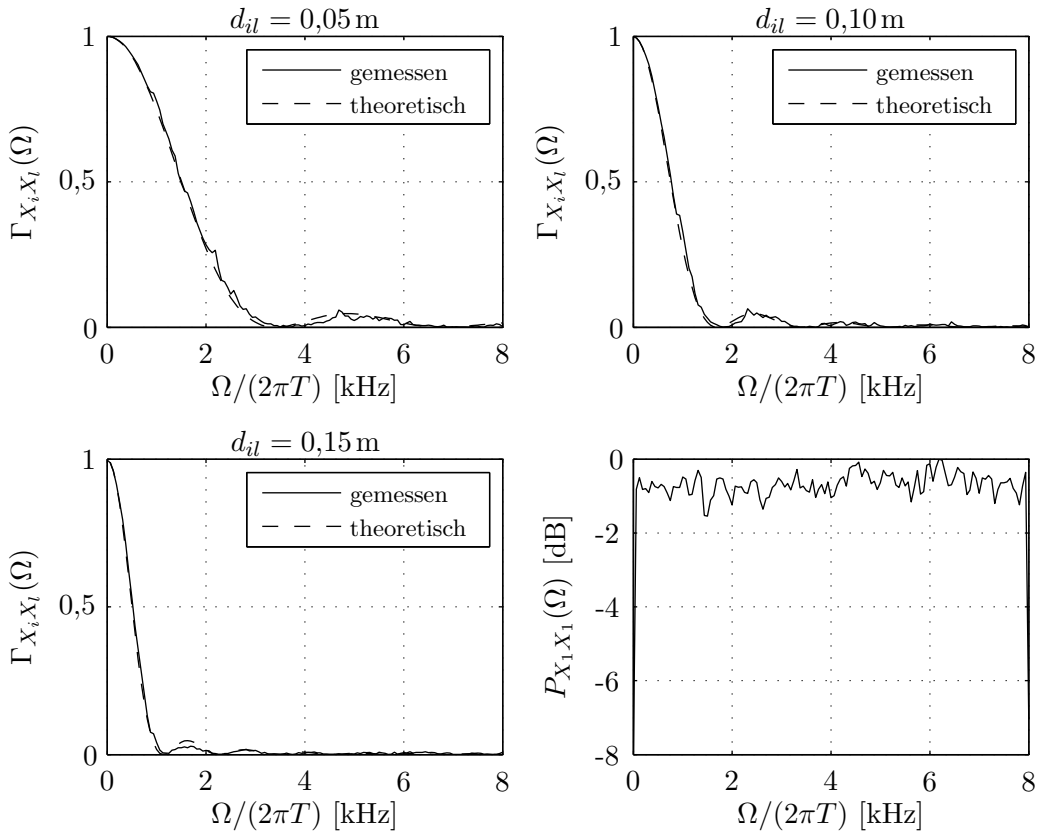

Bild 2.7: Kohärenzverlauf eines simulierten, diffusen Schallfeldes für unterschiedliche Sensorabstände im Vergleich zum idealen Verlauf und das Leistungsdichtespektrum eines Signals.

Aufgrund der zeitvarianten statistischen Eigenschaften von Sprachsignalen werden die statistischen Kenngrößen über kurze Zeitabschnitte von etwa 8 bis 30 ms bestimmt. Die Kurzzeitspektren der Signale können dann mit Hilfe der Methode von Welch gemessen werden [Wel67]. Dabei werden sich überlappende Segmente der Zeitsignale mit einer Fensterfunktion gewichtet und in den Frequenzbereich transformiert. Für das komplexe Kurzzeit-Kreuzleistungsdichtespektrum des m-ten Segments ergibt sich das so genannte Kreuzperiodogramm

$$P_{X_{i,m}X_{l,m}}(\Omega_k) = \frac{1}{L} X^*_{i,m}(\Omega_k) X_{l,m}(\Omega_k) \tag{2.21}$$

und entsprechend das Kurzzeit-Autoleistungsdichtespektrum, bzw. das Autoperiodogramm

$$P_{X_{i,m}X_{i,m}}(\Omega_k) = \frac{1}{L} |X_{i,m}(\Omega_k)|^2. \tag{2.22}$$

Hierbei wird mit $(\cdot)^*$ das konjugiert komplexe Spektrum gekennzeichnet und L gibt wieder die Anzahl der Stützstellen der DFT an. Unter Verwendung des Langzeitmittelwerts der

[10] Für die genaue Unterscheidung zwischen der Definition und dem Messen statistischer Kenngrößen sei z. B. auf [VHH98], [KK02] verwiesen.

Periodogramme erhält man so einen Messwert für die Kohärenz

$$\hat{\Gamma}_{X_i X_l}(\Omega_k) = \frac{\sum_{m=0}^{N-1} |P_{X_{i,m} X_{l,m}}(\Omega_k)|^2}{\sum_{m=0}^{N-1} P_{X_{i,m} X_{i,m}}(\Omega_k) P_{X_{l,m} X_{l,m}}(\Omega_k)}. \tag{2.23}$$

Für eine möglichst zuverlässige Bestimmung der Kohärenz sollte die Anzahl der berücksichtigten Segmente N hinreichend groß gewählt werden. Für den Extremfall von nur einem betrachteten Segment ist zu sehen, dass die Kurzzeit-Kohärenz für alle Frequenzen den Wert Eins annimmt. Einen ebenfalls wichtigen Parameter stellt die Frequenzauflösung

$$\Delta f = \frac{f_{Ab}}{L} \tag{2.24}$$

der zugrundeliegenden diskreten Fouriertransformation dar. In z. B. [JN87], [Mar95], [Dre99] wurde der Zusammenhang zwischen der gemessenen Kohärenz und der Frequenzauflösung untersucht. Die dabei erzielten Ergebnisse führen zu folgender Schlussfolgerung: Bei einer Frequenzauflösung von $\Delta f \gg 4/T_{60}$ wird in einem Schallfeld, bei dem die Mikrophone außerhalb des Hallradius der Schallquellen liegen, stets eine Kohärenz gemessen, die näherungsweise dem si^2-Verlauf nach Gl. (2.20) entspricht. Der korrekte Kohärenzverlauf stellt sich hingegen erst bei $\Delta f < 4/T_{60}$ ein. Für eine übliche Nachhallzeit von 0,4 s in einem Büroraum würde sich so eine notwendige Frequenzauflösung $\Delta f < 10$ Hz ergeben. Da aber bei der Sprachsignalverarbeitung in der Regel Auflösungen zwischen 16 kHz/256 = 62, 5 Hz und 16 kHz/1024 = 15, 625 Hz verwendet werden, "sehen" die Mikrophone nicht die korrekte Kohärenz, sondern einen zur si^2-Funktion ähnlichen Kohärenzverlauf.

Zur Analyse der räumlichen Kohärenz realer Schallfelder wurde eine Reihe von Messungen in unterschiedlichen Räumen durchgeführt. Für die Aufnahmen kamen vier äquidistant im Abstand von 5 cm angeordnete Grenzflächenmikrophone mit Hypernierencharakteristik (AKG C-400BL) zum Einsatz. Die dabei gemessene Langzeit-Kohärenz nach Gl. (2.23) soll hier beispielhaft für einige Anordnungen vorgestellt werden, wobei jeweils eine Frequenzauflösung von $\Delta f = 62{,}5\,\mathrm{Hz}$, eine Überlappung der Segmente von 50 % und ein Hanning-Fenster gewählt wurde.

Zuerst ist in Bild 2.8 die Kohärenz eines aufgenommenen ca. 12 s langen Sprachsignals für zwei Räume zu sehen. In einem Fall wurden die Mikrophone auf einem Stativ in einem reflexionsarmen Raum (Größe: 4 m x 7 m x 3 m) und im anderen Fall waren die Mikrophone auf einem Monitor in einem Büroraum ($T_{60} \approx 0{,}5\,\mathrm{s}$; Größe: 4 m x 5 m x 3 m) angeordnet. Das Sprachsignal wurde über einen Lautsprecher jeweils in einem Abstand von ca. 0,6 m mittig vor den Mikrophonen ausgegeben, also innerhalb des Hallradius. Im reflexionsarmen Raum sind für alle Frequenzen Kohärenzwerte nahe Eins zu beobachten, die jedoch mit steigender Frequenz und zunehmendem Mikrophonabstand abnehmen. Für den Fall des Büroraumes ist bereits ein stark frequenzselektiver Kohärenzverlauf festzustellen, trotz der mittleren Nachhallzeit und des kleinen Abstands zwischen dem Lautsprecher und den Mikrophonen.

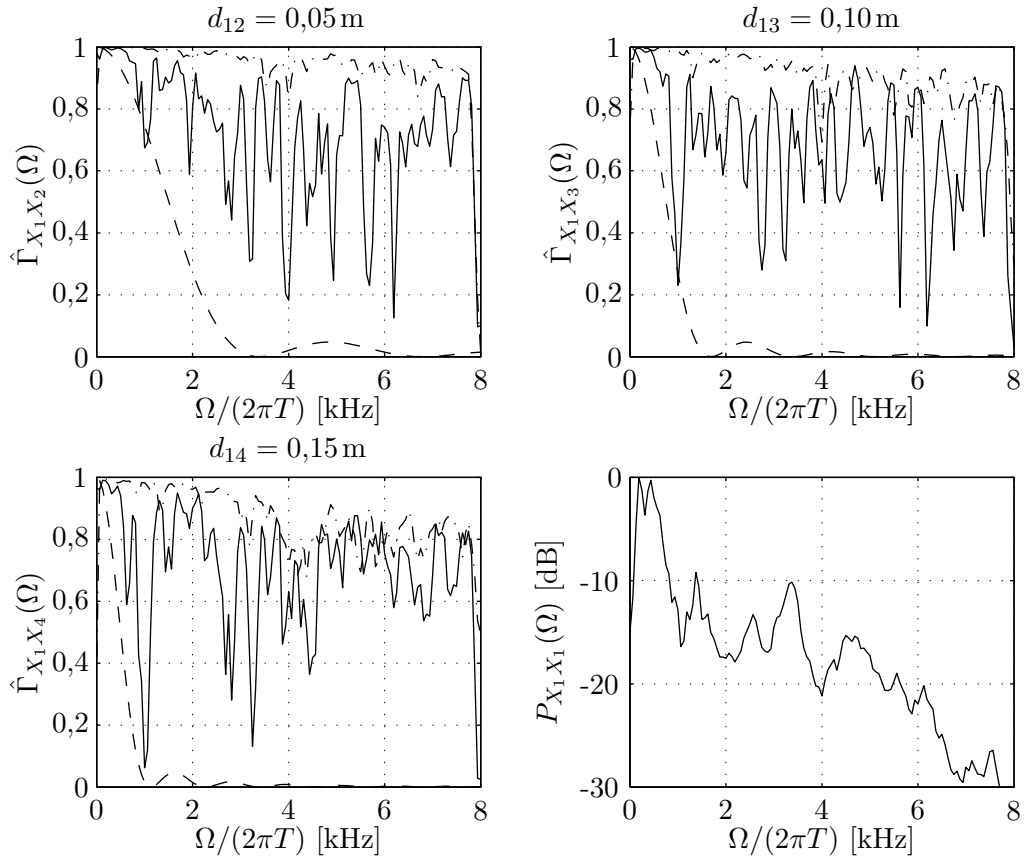

Bild 2.8: Gemessener Kohärenzverlauf eines Sprachsignals im Büroraum (–) und im reflexionsarmen Raum (··) für unterschiedliche Mikrophonabstände bei 0,6 m Abstand zwischen Lautsprecher und *Array* im Vergleich zum si^2-Verlauf (- -), sowie das Autoleistungsdichtespektrum des ersten Signals.

Als nächstes sind in Bild 2.9 die Ergebnisse einer Messung, bei der im Büroraum ein breitbandiges Rauschsignal über einen Lautsprecher in einer Entfernung von 3 m zu den Mikrophonen abgestrahlt wurde. Die Schallquelle befindet sich also außerhalb des Hallradius. Bei den Kohärenzverläufen stellt sich folglich für sehr niedrige Frequenzen ein ähnlicher Verlauf wie bei der si^2-Funktion ein. Abgesehen von den niedrigen Frequenzen ist jedoch eine höhere räumliche Kohärenz zu beobachten als beim diffusen Schallfeld. Der kohärente Schalleinfall des Direktschalls und einiger energiereicher frühen Echos zeigte bei den Messungen durchweg noch große Auswirkungen auf den Kohärenzverlauf, insbesondere bei höheren Frequenzen. Erst bei genügend großen Entfernungen zu den Mikrophonen im Vergleich zum Hallradius ist der Direktschall aufgrund der Ausbreitungsdämpfung soweit abgeklungen, dass er einen sehr geringen Einfluss auf die räumliche Kohärenz am Aufnahmeort nimmt.

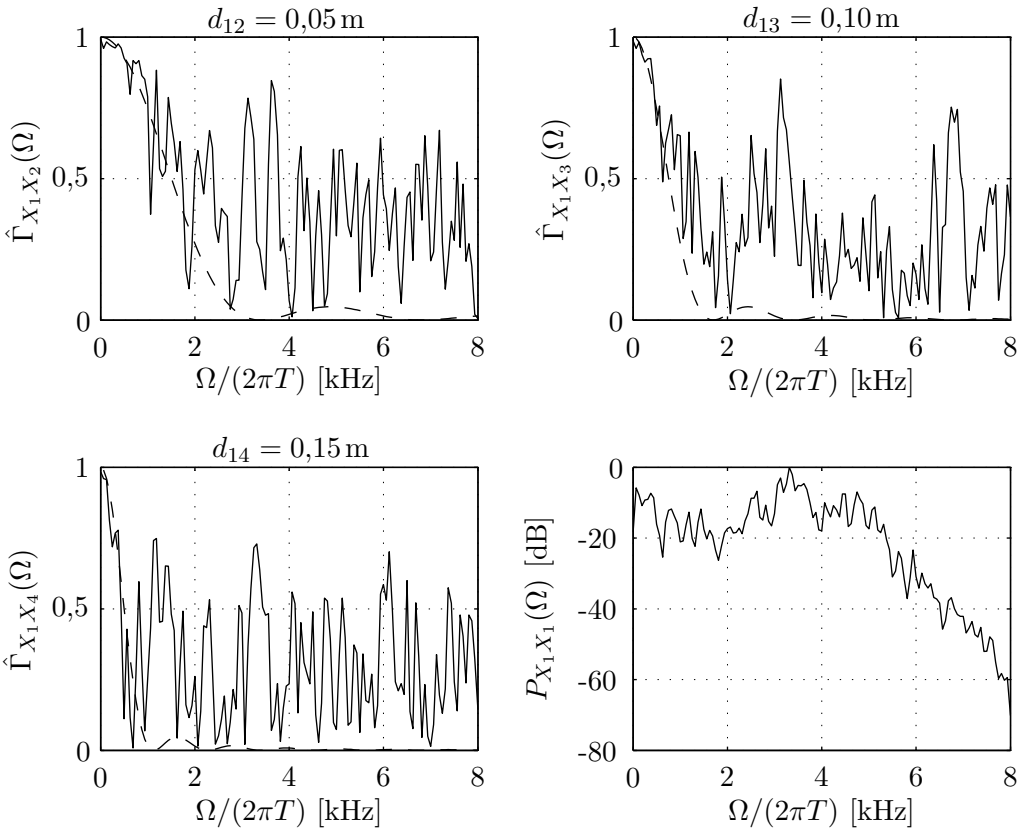

Bild 2.9: Gemessener Kohärenzverlauf eines breitbandigen Rauschsignals im Büroraum (–) für unterschiedliche Mikrophonabstände bei 3 m Abstand zwischen Schallquelle und *Array* im Vergleich zum si^2-Verlauf (- -), sowie Autoleistungsdichtespektrum des ersten Signals.

Um die Auswirkungen des Direktschalls zu untersuchen, wurden verschiedene Anordnungen gewählt, bei denen keine Sichtverbindung zwischen der Schallquelle und den Mikrophonen bestand. Hierbei zeigte sich im Allgemeinen ein zur si^2-Funktion deutlich ähnlicherer Kohärenzverlauf im Vergleich zu Anordnungen mit Direktschallkomponente. In Bild 2.10 ist die Kohärenz für eine Beschallungssituation abgebildet, bei der im Büroraum ein Rechnerlüfter als Schallquelle fungiert hat. Dieser befand sich unter dem Tisch, auf welchem der Monitor mit den Mikrophonen platziert war. An dem gemessenen Autoleistungsdichtespektrum ist der für einen solchen Fall typische Tiefpasscharakter zu erkennen. Trotz der geringen geometrischen Entfernung von ca. 1 m ist die Nachbildung des Hauptmaximums der si^2-Funktion deutlich ausgeprägt. Die Kohärenz verschwindet für hohe Frequenzen aufgrund nicht vorhandener Frequenzkomponenten der Schallquelle, so dass das unkorrelierte Mikrophonrauschen dominiert.

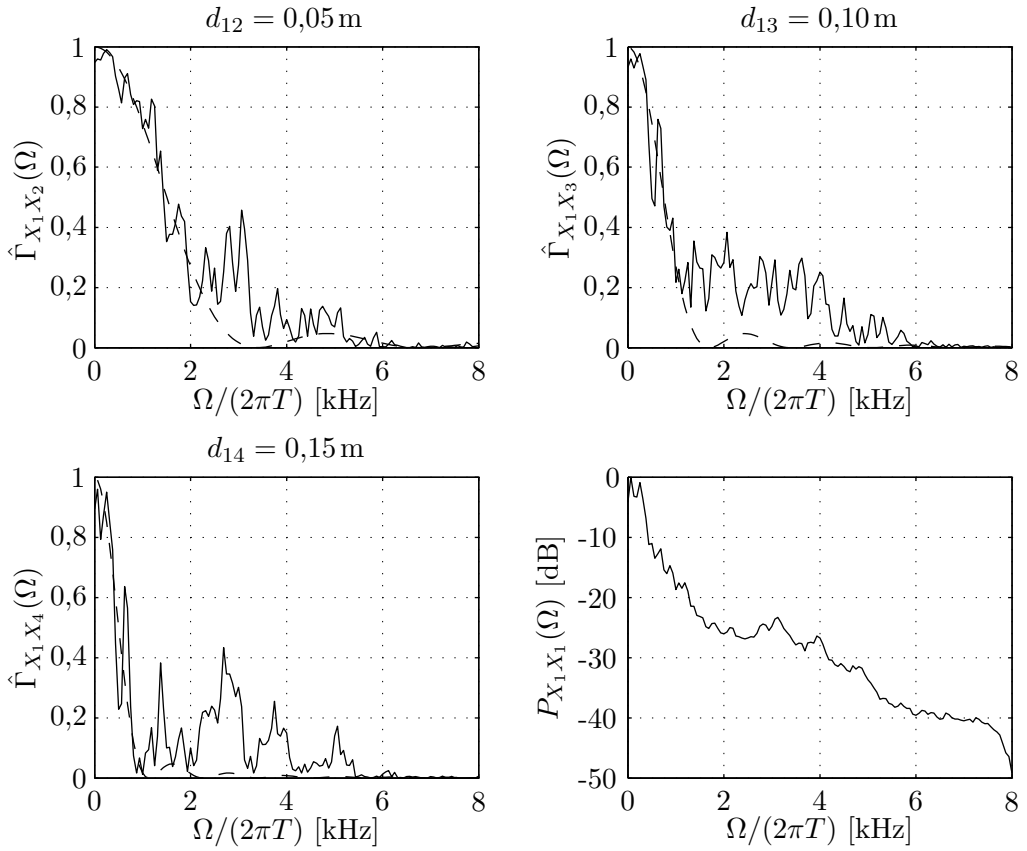

Bild 2.10: Gemessener Kohärenzverlauf im Büroraum mit Rechnerlüfter als indirekte Störquelle (–) für unterschiedliche Mikrophonabstände im Vergleich zum si^2-Verlauf (- -), sowie Autoleistungsdichtespektrum des ersten Signals.

Als letztes sind Messergebnisse zur räumlichen Korrelationseigenschaft eines Laborraums ($T_{60} \approx 0{,}8\,\mathrm{s}$; Größe: 7 m x 10 m x 3 m) in Bild 2.11 dargestellt. Das *Array* befand sich dabei wiederum auf einem Stativ in 1,60 m Höhe. Das Schallfeld wurde durch mehrere Rechnerlüfter, Festplattengeräusche und einen Drucker erzeugt, wobei sich die Rechner jeweils unter Arbeitstischen befanden. Im Vergleich zur Anordnung mit nur einem Rechnerlüfter als Schallquelle im Büroraum liegt der Unterschied also in der Verwendung von mehreren Quellen und einer höheren Nachhallzeit des Raums. Die Folge ist eine deutlich geringere Kohärenz im Bereich der mittleren Frequenzen. Für hohe Frequenzen liegt zwar eine sehr geringe Kohärenz vor, sie verschwindet aber nicht wie in Bild 2.10, da das Szenario im Laborraum ein breiteres Frequenzspektrum aufweist. Aufgrund der höheren Nachhallzeit und einer räumlichen Verteilung der Schallquellen im Abstand von 2 bis 6 m zu den Mikrophonen ist der Anteil an direkten Schallkomponenten sehr gering. Da für die Größe der räumlichen Kohärenz das Verhältnis von Direktschall zu Diffusschall maßgebend ist, lässt sich für diese Anordnung also ein näherungsweise als diffus zu bezeichnendes Schallfeld messen.

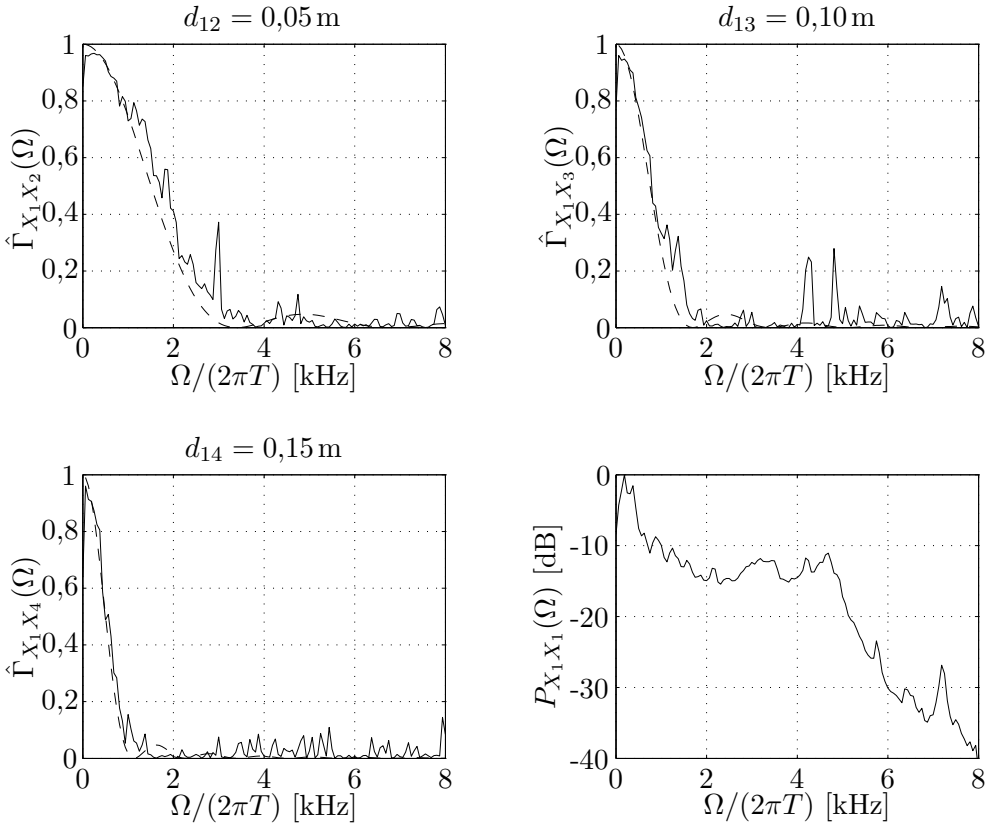

Bild 2.11: Gemessener Kohärenzverlauf im Laborraum mit mehreren Störquellen für unterschiedliche Mikrophonabstände, sowie Autoleistungsdichtespektrum des ersten Signals.

2.5 Zusammenfassung

In diesem Kapitel wurden die wesentlichen Kenngrößen der statistischen Raumakustik eingeführt. Die hier definierten Schallfeldparameter waren die Anfangsnachhallzeit T_A, die Nachhallzeit T_{60}, der Hallradius r_H und das Deutlichkeitsmaß C_{50}. Sie lassen sich mit Hilfe der Raumimpulsantwort bzw. anhand der Rückwärtsintegration der Raumimpulsantwort bestimmen. Bei den späteren experimentellen Untersuchungen der *Beamforming*-Verfahren sollen diese raumakustischen Eigenschaften noch Verwendung finden. Dort werden mit Hilfe der hier vorgestellten Spiegelquellenmethode mehrkanalige akustische Signale erzeugt. Anhand dieser simulierten Sprachdaten ist es möglich die Verfahren für unterschiedliche Anordnungen und Nachhallzeiten zu testen.

Für die Aufnahme von Sprachsignalen mittels Freisprecheinrichtungen kann generell gesagt werden, dass es wünschenswert ist möglichst viel Schallenergie des Nutzsignals aufzunehmen, d. h., der Sprecher sollte sich innerhalb des Hallradius befinden. Aufgrund der in diesem Kapitel gemachten Betrachtungen sind für die Sprachverständlichkeit noch zusätzlich die frühen Reflexionen von Bedeutung.

Weiterhin wurde exemplarisch an beispielhaften Messungen realer Schallfelder die räumliche Kohärenz analysiert. Dabei stellte sich heraus, dass sich insbesondere im Falle von indirekten Schallquellen ein näherungsweise diffuses Schallfeld ergibt. Dies ist bei den späteren

Betrachtungen von Bedeutung, da sich additiv zu einem Sprachsignal häufig Hintergrundrauschen aufgrund indirekter Quellen überlagert, z. B. durch Rechnerlüfter. Werden hingegen Störschallquellen mit direkter Sichtverbindung zu den Mikrophonen platziert, so ist auch bei größeren Abständen zu diesen noch eine deutliche Kohärenz zu messen. Solche Quellen werden daher im Weiteren gesondert betrachtet und als direkte Störschallquellen bezeichnet.

Kapitel 3

Grundlagen zu Mikrophongruppen

Während bei einkanaligen Verfahren zur Sprachsignalverbesserung lediglich spektrale Informationen zur Adaption von zeitvarianten Filtergewichten vorliegen, kann bei der mehrkanaligen Sprachsignalverarbeitung mittels Mikrophongruppen auch die räumliche Komponente der Anordnung genutzt werden. Dabei wird die akustische Welle räumlich abgetastet und mit der anschließenden strahlformenden Signalverarbeitung (engl. *Beamforming*) können Signale aus bestimmten Raumrichtungen gegenüber anderen verstärkt oder unterdrückt werden. Für diese so genannte Raum-Zeit-Filterung kommen üblicherweise Filter mit endlicher Impulsantwort (engl. *Finite Impulse Response*, FIR) in jedem Mikrophonpfad zum Einsatz, wobei die gefilterten und aufsummierten Mikrophonsignale dann das Ausgangssignal des *Beamformers* ergeben. Daher kann solch eine Strahlformung auch allgemein englischsprachig als *Filter-and-Sum-Beamformer* (FSB) bezeichnet werden. In diesem Kapitel soll zunächst der einfachste Fall der Realisierung der FIR-Filter als reine Verzögerungsglieder zur Kompensation der Laufzeitunterschiede der akustischen Welle von der Quelle zu den einzelnen Mikrophonen hin angenommen werden. Es erfolgt eine Beschreibung der Problemstellung bezüglich der Anordnung von Mikrophongruppen und das sich aus dem Aufbau ergebende frequenzabhängige Dämpfungsverhalten. Desweiteren soll zur allgemeinen Bewertung eines Gesamtsystems bestehend aus Mikrophonanordnung und Filterung sowohl auf die objektive Messung von Leistungsmerkmalen wie die Verbesserung des Signal-zu-Rauschverhältnisses, als auch auf subjektive Möglichkeiten zur Beurteilung des verarbeiteten Sprachsignals eingegangen werden. Dabei zeigt sich ein unterschiedliches Verhalten je nach Annahme des vorliegenden Störschallfeldes, welche im vorangegangenen Kapitel eingeführt wurden.

3.1 Beamformer-Signalmodell

Bereits in Abschnitt 2.3 wurde ein mehrkanaliges Signalsystem zur Aufnahme von Störschallfeldern vorgestellt, um die räumliche Kohärenz zu untersuchen. In Bild 3.1 ist diese Anordnung um FIR-Filter in jedem Mikrophonpfad erweitert. Ausgegangen wird wieder von einem Sprecher als Quelle für das Nutzsignal $s_c(t)$ an der Position $\mathbf{p}_s$, einer Störgeräuschquelle $n_c(t)$ positioniert an den Koordinaten $\mathbf{p}_n$ und M Mikrophonen, jeweils bei $\mathbf{p}_i$. Die Positionen $\mathbf{p}_s$, $\mathbf{p}_n$ und $\mathbf{p}_i$, $i \in \{1, 2, \ldots, M\}$ beschreiben vektoriell den jeweiligen Ort im dreidimensionalen Raum. Das Mikrophonsignal erfährt eine Abtastung zu den Zeiten nT, so dass anschließend zeitdiskrete Signale mit der Abtastrate $1/T$ und dem Zeitindex n vorliegen. Zusätzlich

zu den korrelierten zeitdiskreten Störsignalen $\mathbf{n}_c(n)$ sind ebenfalls unkorrelierte Störanteile $n_{u,1}(n), \ldots, n_{u,M}(n)$ in jedem Signalpfad enthalten, welche das Rauschen durch die Mikrophone und die Verstärkung nachbilden. Der im Mikrophonsignal enthaltene Nutzanteil entsteht durch Faltung des Sprachsignals mit den jeweiligen Raumimpulsantworten $h_i(n)$ und der des Rauschanteils aus dem Faltungsprodukt des von der Störquelle abgegebenen Signals mit den Raumimpulsantworten $a_i(n)$. Am *Beamformer*-Ausgang liegt das in jedem Signalpfad gefilterte und dann aufsummierte Signal $y(n)$ vor. In Bild 3.1 sind die FIR-Filter zeitinvers[1]

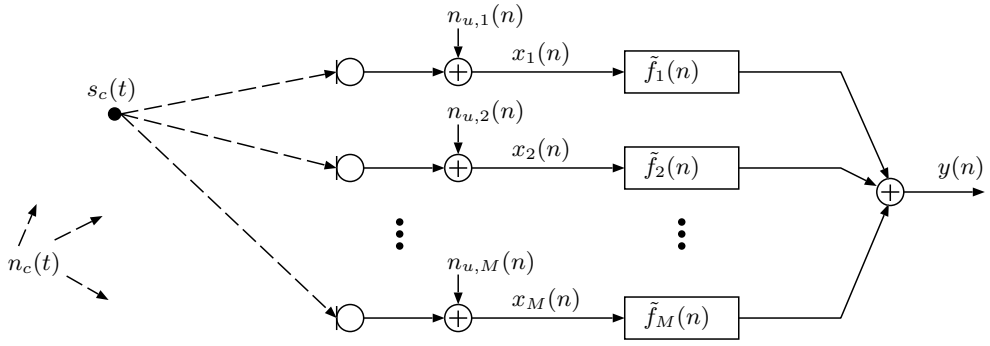

Bild 3.1: Allgemeines Modell eines *Filter-and-Sum-Beamformers*.

mit $\tilde{f}_i(n) = f_i(-n)$ angenommen, so dass der funktionale Zusammenhang für das allgemeine Modell eines *Filter-and-Sum-Beamformers* geschrieben werden kann als

$$y(n) = \sum_{i=1}^{M} \tilde{f}_i(n) * x_i(n), \tag{3.1}$$

wobei das Signal $x_i(n)$ in jedem Signalpfad entsprechend Bild 3.1 wie folgt zusammengesetzt ist:

$$x_i(n) = s_c(n) * h_i(n) + n_c(n) * a_i(n) + n_{u,i}(n) \tag{3.2}$$

$$x_i(n) = s_i(n) + n_i(n). \tag{3.3}$$

In Gl. (3.3) sind Nutz- und Störanteile im i-ten Signalpfad zu

$$s_i(n) = s_c(n) * h_i(n) \tag{3.4}$$

$$n_i(n) = n_c(n) * a_i(n) + n_{u,i}(n) \tag{3.5}$$

zusammengefasst.

Äquivalent zu der bereits erfolgten Beschreibung in Abschnitt 2.3 soll auch hier eine Darstellung aller Signale im Frequenzbereich bevorzugt werden. Die zeitdiskrete Fourier-Transformation von Gl. (3.1) liefert somit folgendes Ergebnis

$$Y(\Omega) = \sum_{i=1}^{M} F_i^*(\Omega) \cdot X_i(\Omega) \tag{3.6}$$

$$= \sum_{i=1}^{M} F_i^*(\Omega) \cdot (S_i(\Omega) + N_i(\Omega)). \tag{3.7}$$

[1]Durch die zeitinverse Notation der FIR-Filter läßt sich die Filterung mittels Vektorschreibweise im Frequenzbereich kompakt durch Gl. (3.13) darstellen.

Weiterhin wird im Folgenden vorzugsweise die Vektornotation der Signale verwendet, deren Komponenten jeweils durch die zugehörigen Signalpfade gegeben sind, z. B. durch

$$\mathbf{H}(\Omega) = [H_1(\Omega), \ldots, H_M(\Omega)]^T \tag{3.8}$$
$$\mathbf{S}(\Omega) = [S_1(\Omega), \ldots, S_M(\Omega)]^T \tag{3.9}$$
$$\mathbf{N}(\Omega) = [N_1(\Omega), \ldots, N_M(\Omega)]^T \tag{3.10}$$
$$\mathbf{X}(\Omega) = [X_1(\Omega), \ldots, X_M(\Omega)]^T \tag{3.11}$$
$$\mathbf{F}(\Omega) = [F_1(\Omega), \ldots, F_M(\Omega)]^T, \tag{3.12}$$

wobei $(\cdot)^T$ die transponierte Schreibweise des jeweiligen Vektors bezeichnet. Mit Hilfe der beschriebenen Vektornotation ergibt sich aus den Gleichungen (3.6) und (3.7)

$$Y(\Omega) = \mathbf{F}^H(\Omega) \cdot \mathbf{X}(\Omega) \tag{3.13}$$
$$= \mathbf{F}^H(\Omega) \cdot (\mathbf{S}(\Omega) + \mathbf{N}(\Omega)) \tag{3.14}$$
$$= \mathbf{F}^H(\Omega) \cdot (S_c(\Omega)\mathbf{H}(\Omega) + \mathbf{N}(\Omega)), \tag{3.15}$$

mit $(\cdot)^H$ für die hermitesch konjugierte Notation. An dieser Stelle soll angemerkt werden, dass alle eingeführten Signale in Gl. (3.8) bis Gl. (3.12) von der konkreten Positionierung der Schallquellen (Positionen $\mathbf{p}_s$ und $\mathbf{p}_n$) und der Mikrophone (Positionen $\mathbf{p}_i$) im Raum abhängen, also nicht nur von der relativen Ausrichtung zueinander, sondern der absoluten Anordnung im Raum. Daher müssten konsequenterweise jeweils diese geometrischen Informationen ebenfalls als Argument der Signale auftreten. Aufgrund einer kürzeren Schreibweise soll auf diese Notation verzichtet werden, so dass z. B. für die Raumübertragungsfunktion folgende Definition gilt:

$$\mathbf{H}(\Omega) := \mathbf{H}(\Omega, \mathbf{p}_s, \mathbf{p}_1, ..., \mathbf{p}_M). \tag{3.16}$$

Äquivalent zur Definition Gl. (3.16) gelten ebenfalls verkürzte Schreibweisen für die Signale in Gl. (3.9) bis Gl. (3.12).

Das Ziel des *Beamformings* ist es nun, die Filterkoeffizienten $\mathbf{F}(\Omega)$ so zu wählen, dass das Quellsignal des Sprechers möglichst gut rekonstruiert wird. Dabei läßt sich der Filter-Entwurf grundsätzlich in zwei unterschiedliche Klassen aufteilen: datenunabhängige und datenabhängige Verfahren [VVB88].

Data-Independent-Beamforming Bei einem datenunabhängigen (engl. *Data-Independent*) *Beamforming*-Verfahren hängen die Filterkoeffizienten nicht von den Eingangsdaten, also den Mikrophonsignalen, ab. Die Filtergewichte werden entsprechend einer gewünschten Raum-Zeit-Übertragungsfunktion entworfen, wobei häufig ein Signal aus einer vorgegebenen Richtung am *Beamformer*-Ausgang erhalten bleiben soll, und weiterhin ein Filter-Design bezüglich der Breite der Hauptkeule und der Höhe der Nebenkeulen erfolgt. Die verschiedenen Formen der *Array*-Gewichtung sind häufig äquivalent zu Fensterfunktionen in der Spektralanalyse. Die eingesetzte spektrale Gewichtung ermöglicht dann die Richtcharakteristik so zu optimieren, dass z. B. die Höhe der Nebenkeulen minimiert wird, oder über alle Frequenzen gemittelt eine Mindestdämpfung der Nebenkeulen erreicht wird; siehe [VT02] für einen Überblick.

Data-Dependent-Beamforming Einem *Beamforming*-Design, welches datenabhängig (engl. *Data-Dependent*) ausgelegt sein soll, liegt die Idee zugrunde, eine zeitvariante Raum-

Zeit-Übertragungsfunktion zu ermöglichen. So kann z. B. auch für zeitlich variierende Sprecherpositionen ein optimales *Beamforming* im Sinne des Entwurfskriteriums durch adaptive Verfahren realisiert werden. In Kapitel 4 werden basierend auf den statistischen Eigenschaften der Mikrophonsignale einige optimale, datenabhängige *Beamforming*-Designs vorgestellt.

Second Order Statistics Zum Entwurf statistisch optimaler *Beamforming*-Verfahren (siehe Kapitel 4) aber auch zur Bewertung eines *Beamforming*-Designs ist es notwendig, statistische Eigenschaften zweiter Ordnung (engl. *Second Order Statistics*) zu betrachten, also Signalleistungen bzw. spektrale Leistungsdichten. Da es sich bei Mikrophonsignalen im Allgemeinen um mittelwertfreie Signale handelt, ist das frequenzabhängige Leistungsdichtespektrum (LDS) $\phi_{YY}(\Omega)$ des *Beamformer*-Ausgangssignals Gl. (3.13) gegeben durch

$$\begin{aligned} \phi_{YY}(\Omega) &= E\{|Y(\Omega)|^2\} && (3.17)\\ &= E\{\mathbf{F}^H(\Omega)\mathbf{X}(\Omega)\mathbf{X}^H(\Omega)\mathbf{F}(\Omega)\} && (3.18)\\ &= \mathbf{F}^H(\Omega)E\{\mathbf{X}(\Omega)\mathbf{X}^H(\Omega)\}\mathbf{F}(\Omega), && (3.19) \end{aligned}$$

wobei $E\{\cdot\}$ den Erwartungswert bezüglich aller Realisierungen der entsprechenden Zufallsvariablen bezeichnet. Unter der Annahme zumindest schwach stationärer[2] Eingangssignale, sowie unkorrelierten Rausch- und Sprachanteilen, kann Gl. (3.19) angegeben werden als

$$\begin{aligned} \phi_{YY}(\Omega) &= \mathbf{F}^H(\Omega)\mathbf{\Phi_{XX}}(\Omega)\mathbf{F}(\Omega) && (3.20)\\ &= \mathbf{F}^H(k)\big[\mathbf{\Phi_{SS}}(\Omega) + \mathbf{\Phi_{NN}}(\Omega)\big]\mathbf{F}(\Omega), && (3.21) \end{aligned}$$

wobei $\mathbf{\Phi_{XX}}(\Omega)$ die Matrix der spektralen Kreuzleistungsdichten der Mikrophonsignale und $\mathbf{\Phi_{SS}}(\Omega)$ bzw. $\mathbf{\Phi_{NN}}(\Omega)$ die Matrizen der spektralen Kreuzleistungsdichten des Sprach- bzw. Rauschanteils sind.

Schmalband Annahme Die Realisierung von *Beamforming*-Verfahren im Frequenzbereich und die Berechnung der Kreuzleistungsdichtespektren erfordert eine Dekomposition des breitbandigen Audiosignals in einzelne Spektralkomponenten und deren unabhängige Verarbeitung voneinander. Dabei wird im gesamten Verlauf dieser Arbeit davon ausgegangen, dass in guter Näherung von der Schmalband Annahme ausgegangen werden kann. Betrachtet man ein Mikrophonsignal, welches als mittelwertfrei und zumindest schwach stationär angenommen wird, und integriert das Leistungsdichtespektrum über einen bestimmten Frequenzbereich $[\Omega_0 - \Delta\Omega, \Omega_0 + \Delta\Omega]$ mit der Mittenfrequenz Ω_0, so ist das Ergebnis proportional der mittleren Leistung des Prozesses in diesem Bereich. Für genügend klein gewählte Bereiche $\Delta\Omega$ soll nun für die Schmalband Annahme[3] $\phi_{X_iX_i}(\Omega)$ als näherungsweise konstant innerhalb des betrachteten Intervalls gelten [HN76, VVB88]:

$$\int_{\Omega_0-\Delta\Omega}^{\Omega_0+\Delta\Omega} \phi_{X_iX_i}(\Omega)d\Omega \approx 2 \cdot \Delta\Omega \cdot \phi_{X_iX_i}(\Omega_0). \qquad (3.22)$$

[2]Ein stochastischer Prozess ist stark stationär, wenn dessen Verteilung unabhängig von dem absoluten Zeitindex ist. Hingegen ist ein stochastischer Prozess schwach stationär, wenn lediglich der Erwartungswert unabhängig von dem absoluten Zeitindex ist.

[3]Für die Schmalband Annahme ist es notwendig, dass die Spektralkomponenten untereinander unkorreliert sind. Dies gilt jedoch nur asymptotisch für unendlich lange Beobachtungsfenster [HN76].

3.2 Delay-and-Sum-Beamformer

Die einfachste Form der Realisierung der Filterkoeffizienten $\mathbf{F}(\Omega)$ besteht darin, gerade die Laufzeitdifferenzen für die direkten Ausbreitungspfade der akustischen Welle zwischen der Quelle und den einzelnen Mikrophonen zu kompensieren, um die einzelnen Signale anschließend kohärent zu addieren. Dabei ist zusätzlich auf unterschiedliche Signaldämfungen in den einzelnen Mikrophonpfaden zu achten. Diese entstehen einerseits durch die unterschiedliche Dämpfung aufgrund verschieden langer Ausbreitungspfade, und andererseits durch eine ungleiche Verstärkung der Mikrophonsignale bzw. uneinheitliche Mikrophoncharakteristiken. Solch eine Strahlformung, die lediglich aus den Verzögerungen, einer reellwertigen, skalaren Gewichtung und der anschließenden Summation besteht, wird *Delay-and-Sum-Beamformer* (DSB) genannt. Unter der idealen Annahme, dass die beschriebene Dämpfung in jedem Pfad identisch ist, reduziert sich die Gewichtung auf $1/M$ um den Signalpegel des Nutzsignals vom Eingang zum Ausgang bei M kohärent addierten Signalen konstant zu halten. Im Weiteren soll nun dieser Sachverhalt formal beschrieben und wichtige Begriffe eingeführt werden.

Beamformer Response Es soll nun angenommen werden, dass $\mathbf{p}_s$ die Position einer monochromatischen Quelle

$$s_c(n) = S_c \cdot e^{j\Omega n} \tag{3.23}$$

der normierten Frequenz Ω mit der Amplitude S_c angibt. Die Laufzeit des Signals von der Quelle bis zum i-ten Mikrophon an der Stelle $\mathbf{p}_i$ ist dann

$$\tau_i := \tau_i(\mathbf{p}_s, \mathbf{p}_i) = \frac{1}{c}||\mathbf{p}_s - \mathbf{p}_i||. \tag{3.24}$$

Das Quellsignal $s_c(n)$ gelange ohne Reflexionen und Dämpfung zu den Mikrophonen, wo sich jeweils das Signal

$$s_i(n) = S_c e^{j\Omega(n-\tau_i/T)} \tag{3.25}$$

ergibt. Das Signal am *Beamformer*-Ausgang kann dann entsprechend Gl. (3.7) geschrieben werden als

$$y(n) = \sum_{i=1}^{M} F_i^*(\Omega) \cdot S_c e^{j\Omega(n-\tau_i/T)}. \tag{3.26}$$

Aus Gl. (3.26) kann somit die Antwort des *Beamformers* (engl. *Beamformer Response*) auf ein von der Position $\mathbf{p}$ auf die Sensorgruppe einfallendes Signal entsprechend [VVB88] definiert werden:

$$r(\Omega, \mathbf{p}) := \sum_{i=1}^{M} F_i^*(\Omega) \cdot e^{-j\Omega||\mathbf{p}-\mathbf{p}_i||/(Tc)}. \tag{3.27}$$

Am *Beamformer*-Ausgang ergibt sich dann in kompakter Schreibweise

$$y(n) = S_c e^{j\Omega n} \cdot r(\Omega, \mathbf{p} = \mathbf{p}_s). \tag{3.28}$$

Steering Vector Möchte man nun wie eingangs beschrieben eine Laufzeitkompensation in jedem Signalpfad realisieren, sind äquivalent zu Gl. (3.27) Exponentialterme einzuführen. Hier nun allerdings aus Sicht des *Arrays*, d. h. durch geeignete Verzögerungen kann die "Blickrichtung" (engl. *Look Direction*) des *Arrays* auf ein Ziel (engl. *target*) $\mathbf{p}_t$ hin ausgerichtet werden. Die Zielkoordinaten $\mathbf{p}_t$ sollten dabei idealerweise gleich den Quellkoordinaten sein $\mathbf{p}_t = \mathbf{p}_s$,

bzw. einer möglichst guten Schätzung dieser entsprechen. Die Laufzeitdifferenz, welche bei einer Ausrichtung auf ein gewünschtes Ziel auszugleichen ist, ergibt sich dann äquivalent zu Gl. (3.24) durch

$$\tau_i(\mathbf{p}_t) := \tau_i(\mathbf{p}_t, \mathbf{p}_i) = \frac{1}{c}||\mathbf{p}_t - \mathbf{p}_i||, \tag{3.29}$$

so dass sich die Exponentialterme als *Steering Vector*[4]

$$\mathbf{d}(\Omega, \mathbf{p}_t) = (e^{j\Omega\tau_1(\mathbf{p}_t)/T},\ e^{j\Omega\tau_2(\mathbf{p}_t)/T}, \ldots, e^{j\Omega\tau_M(\mathbf{p}_t)/T})^H. \tag{3.30}$$

schreiben lassen. Zu beachten ist in Gl. (3.29), Gl. (3.30) und den folgenden Gleichungen, dass die Zielrichtung $\mathbf{p}_t$ als Argument beibehalten wird. Dies ist aus dem Grunde wichtig, da die Ausrichtung des *Arrays* nicht zwangsläufig mit den Quellkoordinaten des Sprechers übereinstimmen müssen.

Grundsätzlich ist es nicht notwendig, die absoluten Laufzeitdifferenzen zwischen der Schallquelle und den Sensoren auszugleichen, sondern lediglich die relativen Zeitdifferenzen bezogen auf einen frei gewählten Raumpunkt wie z. B. den Mittelpunkt der Mikrophongruppe. Die Realisierung von Verzögerungseinheiten, die nicht in das Abtastintervall fallen, kann durch so genannte *Fractional Delay Filter* mit kleinen Approximationsfehlern erfolgen [LVKL96]. Hier soll allerdings der Einfachheit halber die Form in Gl. (3.30) beibehalten werden.

Uniformly Weighted Beamformer Ausgehend von dem *Steering Vector* Gl. (3.30) ist schließlich noch eine einheitliche Gewichtung der *Beamformer*-Signalpfade (engl. *Uniformly Weighted Beamformer*) mit $1/M$ durchzuführen. Die Filterkoeffizienten des idealen *Delay-and-Sum-Beamformers*

$$\mathbf{F}_{\mathrm{DSB}}(\Omega) = \frac{1}{M}\mathbf{d}(\Omega, \mathbf{p}_t = \mathbf{p}_s), \tag{3.31}$$

erzeugen dann am *Beamformer*-Ausgang das Signal

$$Y(\Omega) = \mathbf{F}^H_{\mathrm{DSB}}(\Omega)\mathbf{X}(\Omega). \tag{3.32}$$

Es kann leicht geprüft werden, dass mit Gl. (3.32) das monochromatische Eingangssignal Gl. (3.23) am Ausgang des *Delay-and-Sum-Beamformers* exakt rekonstruiert wird.

Häufig wird in der Literatur die Laufzeitkompensation als *Beamsteering* bezeichnet und als Vorverarbeitungsstufe für das "eigentliche *Beamforming*" durchgeführt. D. h. also, dass für das *Beamforming*-Design von einem mehrkanaligen, so genannten *Presteered*-Signal ausgegangen wird. Obschon in solch einer Anordnung die Laufzeitsteuerung adaptiv auf mögliche Sprecherbewegungen ausgelegt sein kann, sei hier noch angemerkt, dass bei einer datenunabhängigen, fest eingestellten nachfolgenden spektralen Gewichtung auch häufig von einem *Fixed Beamformer* gesprochen wird. Das in dieser Arbeit vorgestellte *Beamformig*-Konzept soll jedoch gerade ohne *a priori* Wissen bezüglich der Sprecherrichtung auskommen, weshalb die Laufzeitkompensation nicht als abgekoppelte Einheit betrachtet werden soll.

3.3 Anordnung der Mikrophone

Die wohl wichtigste Anordnung von Mikrophonen innerhalb einer Gruppe, die insbesondere bei einer geringen Anzahl von Mikrophonen häufig gewählt wird, ist eine äquidistante

[4] Da eine elektronische und nicht physikalische Ausrichtung des *Arrays* gemeint ist, wird auch manchmal statt *Steering Vector* der Begriff *Phase Steering* benutzt.

Platzierung der Mikrophone zueinander. In Bild 3.2 ist solch ein lineares *Array* mit vier Mikrophonen und dem Abstand d zueinander dargestellt[5]. Weiterhin ist in dem Bild die *Broadside*-Blickrichtung (senkrecht zum *Array*), die *Endfire*-Blickrichtung (entlang der Verbindungsachse der Mikrophone) und eine Wellenfront für eine beliebige Einfallsrichtung θ relativ zur *Broadside*-Blickrichtung zwecks Definition der Begriffe eingetragen. Unter der

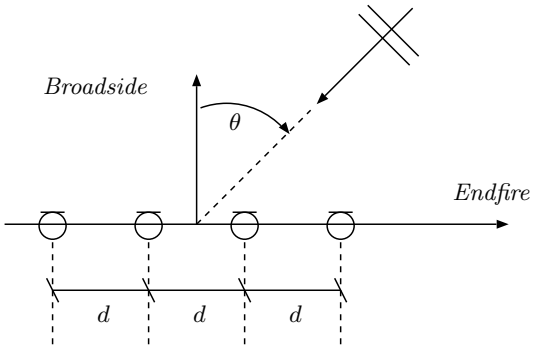

Bild 3.2: Lineare Anordnung einer Mikrophongruppe.

Fernfeld-Annahme, dass also der Schall als planare Welle auf die Mikrophone trifft, "sieht" jedes Mikrophon die Quelle aus der gleichen Richtung[6]: $\theta_i = \theta_t,\ i = 1, \ldots, M$. Es ergibt sich so für die Verzögerung des i-ten Mikrophonsignals bezüglich des *Array*-Mittelpunkts als Referenz

$$\tau_i(\mathbf{p}_t) = \tau_i(\theta_t) = \left(\frac{M+1}{2} - i\right)\frac{d\sin(\theta_t)}{c} \tag{3.33}$$

und folglich für den *Steering Vector* aus Gl. (3.30)

$$\mathbf{d}(\Omega, \mathbf{p}_t) = \mathbf{d}(\Omega, \theta_t) = (e^{j\Omega\tau_1(\theta_t)/T},\ e^{j\Omega\tau_2(\theta_t)/T}, \ldots, e^{j\Omega\tau_M(\theta_t)/T})^H. \tag{3.34}$$

Bei der linearen Anordnung nach Bild 3.2 stellt sich nun die Frage nach einer geeigneten Wahl für den Mikrophonabstand d. Unter der praktisch relevanten Annahme, dass dem Sprachsignal ein diffuses Störschallfeld überlagert ist, kann aus den Betrachtungen der räumlichen Kohärenz im vorangegangenen Kapitel folgendes gesagt werden: Einerseits ist es notwendig die Mikrophone möglichst weit auseinender zu platzieren um eine geringe Kreuzkorrelation für das Störschallfeld zu erhalten und dieses somit in der nachfolgenden Signalverarbeitung gut zu unterdrücken. Andererseits sollte ein kleiner Abstand der Mikrophone gewählt werden, damit das Sprachsignal über den gesamten Frequenzbereich eine hohe Kreuzkorrelation aufweist. Da jedoch davon ausgegangen werden kann, dass sich der Sprecher in einer geringen Distanz zum *Array*, also innerhalb des Hallradius befindet, ist ebenfalls bei größeren Mikrophonabständen noch eine starke Kreuzkorrelation auch bei höheren Frequenzen zu erwarten (siehe gemessenen Kohärenzverlauf eines Sprachsignals in Bild 2.8).

Ein weiteres, entscheidendes Kriterium bezüglich der Wahl des Mikrophonabstandes ist die Mehrdeutigkeit (engl. *Aliasing*) bei der räumlichen Abtastung der akustischen Welle. Um dieses räumliche *Aliasing* auszuschließen, darf der Abstand zwischen den Mikrophonen höchstens der halben minimalen Wellenlänge $\lambda_{\min}$, welche im Wellenfeld auftritt, betragen.

[5] Häufig sind lineare Mikrophongruppen entlang der z-Achse im kartesischen Koordinatensystem angeordnet. Der Zusammenhang zwischen den kartesischen Koordinaten (x, y, z) und den Kugelkoordinaten (r, θ, φ) ist im Anhang in Bild B.1 zu finden.

[6] Die Berechnung des Einfallswinkels einer sphärischen Wellenfront kann in [JD93] gefunden werden.

Für zeitdiskrete Signale korrespondiert die minimale Wellenlänge zur Abtastrate des Systems, so dass sich für den Abstand

$$d \leq \frac{\lambda_{\min}}{2} = Tc \tag{3.35}$$

ergibt. In der Literatur ist häufig für die mehrkanalige Sprachsignalverarbeitung eine Abtastrate von $1/T = f_{Ab} = 8\,\text{kHz}$ zu finden. Da dabei jedoch nur Frequenzen von maximal 4 kHz berücksichtigt werden, klingt das verarbeitete Signal oftmals etwas dumpf, weshalb im Verlauf dieser Arbeit höhere Abtastraten zum Einsatz kommen. Für eine Abtastrate von beispielsweise $f_{Ab} = 12\,\text{kHz}$ ergibt sich dann ein maximaler Mikrophonabstand von 2,83 cm bei einer Schallgeschwindigkeit von $c = 340\,\text{m/s}$. Um die Auswirkung des Mikrophonabstands und der Anzahl der verwendeten Mikrophone zu untersuchen, soll die Richtcharakteristik des *Arrays* analysiert werden.

Beampattern Die Richtcharakteristik (engl. *Beampattern*) ergibt sich aus der Auswertung der *Beamformer Response* in Gl. (3.27) für alle Raumrichtungen. Da hier allerdings nur lineare *Arrays* betrachtet werden, ist das *Beampattern* rotationssymmetrisch und somit unabhängig von der Elevation φ:

$$B(\Omega, \theta) = B(\Omega, \theta, \varphi) = r(\Omega, \mathbf{p}). \tag{3.36}$$

Das *Beampattern* $B(\Omega, \theta)$ wird also im Folgenden verstanden als räumliche Übertragungsfunktion des *Beamformers* (*Beamformer Response*) auf eine planar einfallende Schallwelle aus der Raum-Richtung $\theta = [-\pi/2; \pi/2]$ in Abhängigkeit von der Frequenz.

Im Falle des *Uniformly Weighted Delay-and-Sum-Beamformers* ergeben sich einfach zu analysierende Eigenschaften bezüglich der Richtcharakteristik. Mit Gl. (3.34) ergibt sich der Koeffizientenvektor

$$\mathbf{F}_{\text{DSB}}(\Omega) = \frac{1}{M}\mathbf{d}(\Omega, \theta_t) \tag{3.37}$$

und schließlich das *Beampattern*

$$B_{\text{DSB}}(\Omega, \theta) = \frac{1}{M}\mathbf{d}^H(\Omega, \theta_t)\mathbf{d}(\Omega, \theta) \tag{3.38}$$

$$= \frac{1}{M}\sum_{i=1}^{M} e^{j\Omega\left(\frac{M+1}{2}-i\right)\tau_e/T} \tag{3.39}$$

mit der effektiven Verzögerung

$$\tau_e = \frac{d}{c}\left(\sin(\theta_t) - \sin(\theta)\right) \tag{3.40}$$

bezüglich des *Array*-Mittelpunkts. Mit Hilfe der Formel für die geometrische Reihe kann Gl. (3.38) umgeformt werden zu

$$B_{\text{DSB}}(\Omega, \theta) = \frac{1}{M}\frac{\sin\left(M\Omega\frac{\tau_e}{2T}\right)}{\sin\left(\Omega\frac{\tau_e}{2T}\right)}. \tag{3.41}$$

Anhand der grafischen Darstellung der Richtcharakteristik kann das Prinzip des *Beamformings* verdeutlicht werden: durch Gleichung Gl. (3.41) kann das *Beampattern* entweder für feste Werte von τ_e über die Frequenz oder für feste Frequenzen Ω über den Winkel θ bei eingestellter Ausrichtung θ_t aufgetragen werden. In Bild 3.3 ist das *Beampattern* beispielhaft für

die *Endfire*-Blickrichtung $\theta = \pi/2$ bei gegebener Zielrichtung $\theta_t = 0$ und einer Anordnung aus $M = 5$ Mikrophonen über der auf die Geometrie normierten Frequenz $f \cdot d/c = \Omega d/(2\pi Tc)$ logarithmisch dargestellt[7]. An Bild 3.3 ist die Periodizität des Betrages des *Beampatterns* be-

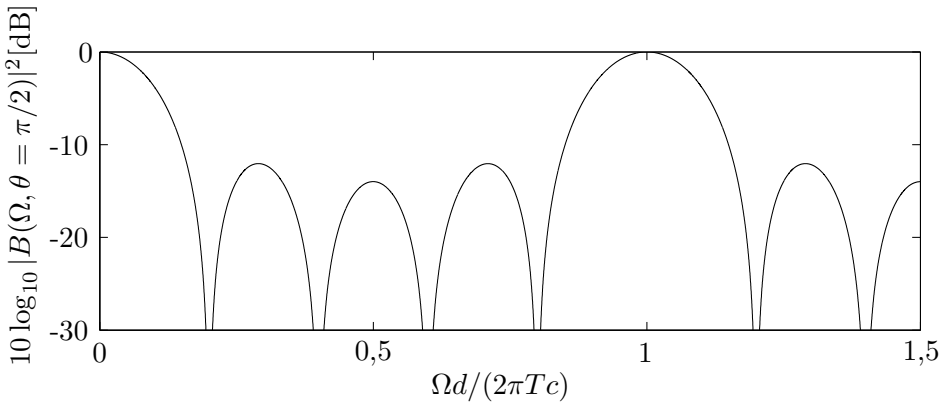

Bild 3.3: Logarithmische Darstellung des DSB-*Beampatterns* über der Frequenz mit $M = 5$ Mikrophonen bei äquidistantem Mikrophonabstand, *Endfire*-Blickrichtung $\theta = \pi/2$ und Zielrichtung $\theta_t = 0$.

züglich Ω mit der Periode[8] $2\pi/\tau_e$ zu erkennen. Gl. (3.41) ist nun weiterhin derart zu interpretieren, dass bei einer Frequenz von 0 Hz von der *Broadside*- bis zur *Endfire*-Richtung, also über den gesamten Winkelbereich, die Übertragungsfunktion konstant ist. Mit steigender Frequenz nimmt die Dämpfung zu den Seiten zu, bis schließlich bei der Frequenz $\Omega d/(2\pi Tc) = 1/M$ die erste Nullstelle und somit die komplette Hauptkeule dargestellt ist. Bei gegebenem Mikrophonabstand kann also mit steigender Mikrophonanzahl auch bei niedrigen Frequenzen eine gute Richtwirkung erreicht werden. Nach der ersten Nullstelle entstehen mit weiter ansteigender Frequenz zusätzlich Nebenkeulen in der Richtcharakteristik.

Betrachtet man nun das Betragsquadrat des *Beampatterns* in Bild 3.4, so ist der beschriebene Sachverhalt in Abhängigkeit des Raumwinkels θ zu beobachten. Dabei ist die Richtcharakteristik logarithmisch oben für die *Broadside*-Blickrichtung $\theta_t = 0°$ und unten für die *Endfire*-Blickrichtung $\theta_t = 90°$, jeweils links für $\Omega d/(2\pi Tc) = 0,1$ und rechts für $\Omega d/(2\pi Tc) = 0,4$ aufgetragen. Zu höheren Frequenzen hin steigt allgemein die Anzahl der Nebenkeulen und die Breite der Hauptkeule nimmt ab. Die Hauptkeule sollte in die Richtung des Sprechers weisen, so dass bei exakter Ausrichtung die Sprachkomponenten synchron und unverzerrt aufsummiert werden. Andererseits bewirkt die ungleichphasige Überlagerung eines kohärenten Schalleinfalls aus anderen Richtungen stets eine Signaldämpfung. Aber auch bei inkohärenten Signalen führt die Mittelung aufgrund der stochastischen Phasenbeziehungen zu einer Signaldämpfung. In Bild 3.4 zeigt sich bei sonst gleichen Werten für d, Ω und M eine unterschiedlich breite Hauptkeule für die *Broadside*- und *Endfire*-Richtung. Die Breite der Hauptkeule ist durch die erste Nullstelle von Gl. (3.41) gegeben, also durch $M\Omega\tau_e/(2T) = \pm\pi$. Für die Richtung der ersten Null des *Beampatterns* gilt dann

$$\sin(\theta) = \sin(\theta_t) \mp \frac{2\pi Tc}{Md\Omega}. \tag{3.42}$$

An Gl. (3.42) ist zu sehen, dass die Breite der Hauptkeule einerseits zu höheren Frequenzen hin und andererseits durch Vergrößerung der Apertur $(M-1)\cdot d$ abnimmt.

[7]Das Betragsquadrat des *Beampatterns* wird auch *Powerpattern* genannt.

[8]Die Funktion in Gl. (3.41) ist bezüglich $\Omega\tau_e/T$ für gerade M 2π-periodisch. Für ungerade M sind die Maxima bei $\pm 2\pi, \pm 6\pi$ negativ und entsprechend bei $\pm 4\pi, \pm 8\pi$ positiv; es liegt eine 4π-Periodizität vor.

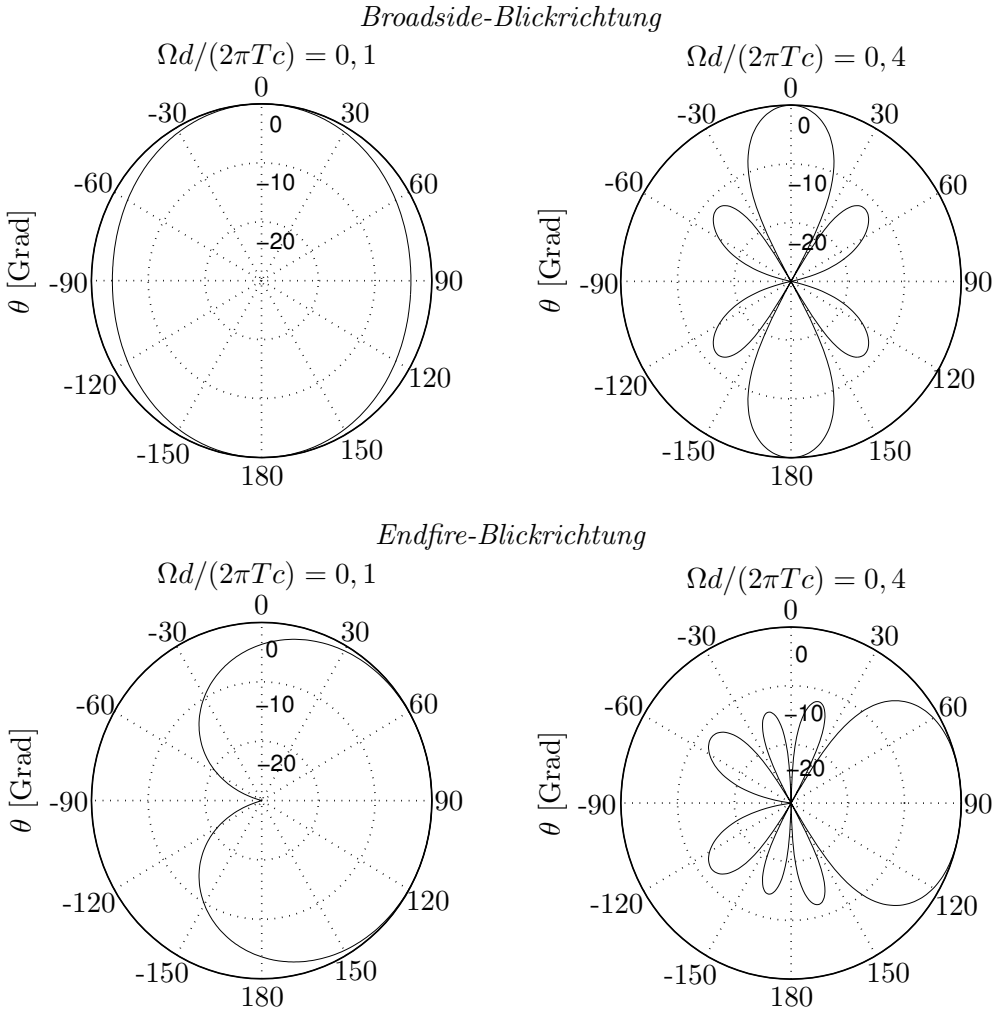

Bild 3.4: Beispielhaftes DSB-*Beampattern* über dem Winkel θ mit $M = 5$ Mikrophonen bei äquidistantem Mikrophonabstand ohne räumliches *Aliasing*. Oben für die Zielrichtung $\theta_t = 0°$ und unten $\theta_t = 90°$; jeweils links für $\Omega d/(2\pi Tc) = 0,1$ und rechts für $\Omega d/(2\pi Tc) = 0,4$.

Wie bereits erwähnt, ist das *Beampattern* $B(\Omega, \theta)$ periodisch in Ω/T mit der Periodendauer $2\pi/\tau_e$, d. h. sie ist abhängig von der Zielrichtung θ_t und der Richtung θ an dem das *Beampattern* ausgewertet wird. Für das räumliche *Aliasing* bedeutet dieser Zusammenhang, dass eine Vieldeutigkeit beim Durchlaufen der Frequenz zuerst bei einer *Endfire*-Ausrichtung $\theta_t = \pm\pi/2$ an der gegenüberliegenden Seite des *Arrays* bei $\theta = \mp\pi/2$ vorliegt. Dann gilt für die effektive Verzögerung $\tau_e = 2d/c$. Nebenkeulen, welche die gleiche Höhe haben wie die Hauptkeule werden *Grating Lobes* genannt. An den Stellen der *Grating Lobes* kann also folglich keine Unterdrückung der Störgeräusche aus den entsprechenden Einfallsrichtungen erfolgen. In Bild 3.5 ist der Effekt des räumlichen *Aliasing* beispielhaft veranschaulicht. Zu sehen ist dort die Richtcharakteristik in der oberen Reihe für die Zielrichtung $\theta_t = 0°$ und unten für $\theta_t = 90°$. Dabei ist jeweils links die normierte Frequenz zu $\Omega d/(2\pi Tc) = 0,5$ und rechts zu $\Omega d/(2\pi Tc) = 1,2$ gewählt.

Um eine weniger stark frequenzabhängige Richtcharakteristik zu erhalten, kann einerseits wie bereits erwähnt eine spektrale Gewichtung als *Fixed Beamformer* mit entsprechender Optimierungsbedingung eingesetzt werden. Eine weitere Möglichkeit beim Einsatz einer grö-

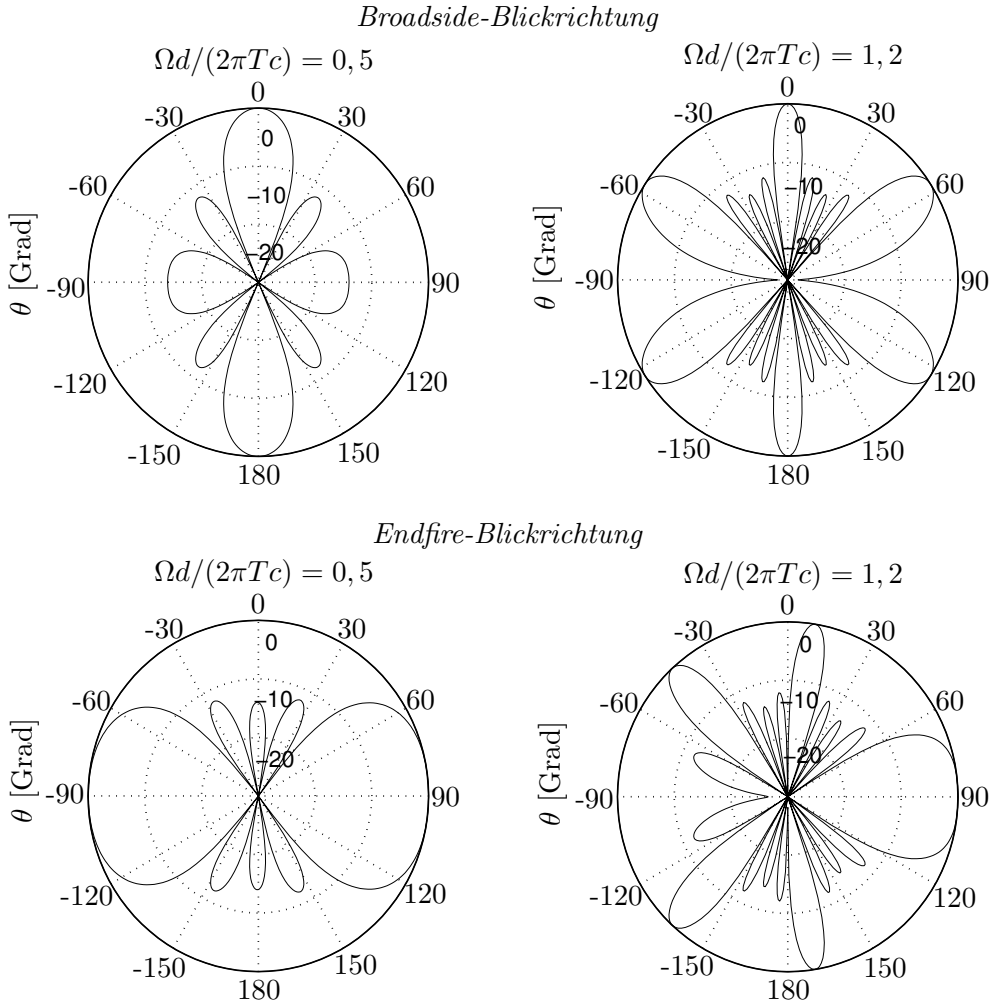

Bild 3.5: Veranschaulichung des räumlichen *Aliasings* für DSB mit $M = 5$ Mikrophonen bei äquidistantem Mikrophonabstand. Oben für die Zielrichtung $\theta_t = 0°$ und unten $\theta_t = 90°$; jeweils links für $\Omega d/(2\pi Tc) = 0,5$ und rechts für $\Omega d/(2\pi Tc) = 1,2$.

ßeren Anzahl an Mikrophonen ist es, diese in Teil-*Arrays* mit jeweils äquidistant zueinander angeordneten Mikrophonen zu nutzen. Dabei ist es möglich einige Mikrophonsignale mehrfach in den unterschiedlichen Teil-*Arrays* auszuwerten, wodurch sich zwischen bestimmten Mikrophonen ein logarithmischer Abstand ergibt [WKW01]. Die Teil-*Arrays* führen dann getrennt für unterschiedliche Frequenzbereiche ein *Beamforming* durch, wobei das Ziel der Mikrophonanordnung und des Filterentwurfs ist, ein möglichst konstantes *Beampattern* über den gesamten, interessierenden Frequenzbereich zu erhalten.

Es sei noch erwähnt, dass neben den linear angeordneten Mikrophonen zweidimensionale Mikrophongruppen von großer Bedeutung sind und in verschiedenen Varianten in der Literatur diskutiert werden. Dabei ist z. B. eine Möglichkeit, die Mikrophone auf der gesamten Fläche eines Rechtecks gleichmäßig zu verteilen. Weitere, häufiger zu findende Anordnungen sind jedoch Kreuz-, Quadratkanten- oder Kreis-Mikrophongruppen [VT02]. Solche, aus vielen Mikrophonen bestehende *Arrays*, sind allerdings eher in der Schallfeldanalyse zu finden, und weniger zur mehrkanaligen Sprachsignalverarbeitung bei Freisprecheinrichtungen, wie es Gegenstand dieser Arbeit sein soll.

Im Weiteren werden ausschließlich lineare Mikrophongruppen eingesetzt mit variierender

Anzahl M, einem Abstand der Mikrophone zueinander von $d = 4\,\text{cm}$ und einer Abtastrate von $f_{Ab} = 12\,\text{kHz}$.

3.4 SNR-basierte Bewertungsgrößen des Beamformings

Zur objektiven Bewertung der erzielbaren Geräuschreduktion von *Beamforming*-Verfahren sind quantitativ messbare Größen wünschenswert. Hier bieten SNR-basierte Methoden ein einfaches Hilfsmittel zur Bestimmung von Kenngrößen, die überdies eine genaue analytische Berechnung zulassen.

Array Gain Eine der wichtigsten objektiv messbaren Bewertungsgrößen bezüglich der Leistungsfähigkeit von *Beamformern* stellt die Verbesserung des Signal-zu-Rauschabstandes vom Eingang zum Ausgang des *Beamformers* dar. Dieser SNR-Gewinn (engl. SNR *Gain*) wird häufig mit dem allgemeinen Ausdruck *Array Gain* bezeichnet [VT02]. Der SNR-Gewinn $G(\Omega)$ wird frequenzabhängig angegeben, da es sich bei Sprache um ein breitbandiges Signal handelt:

$$G(\Omega) = \frac{\text{SNR}_{\text{Array}}(\Omega)}{\text{SNR}_{\text{Sensor}}(\Omega)}, \tag{3.43}$$

wobei mit $\text{SNR}_{\text{Sensor}}(\Omega)$ das frequenzabhängige SNR an den Sensoren und mit $\text{SNR}_{\text{Array}}(\Omega)$ das frequenzabhängige SNR am Ausgang des *Beamformers* bezeichnet ist. Das Signal-zu-Rauschverhältnis des i-ten Sensors ist gegeben durch

$$\text{SNR}_{\text{Sensor},i}(\Omega) = \frac{\phi_{S_iS_i}(\Omega)}{\phi_{N_iN_i}(\Omega)} \tag{3.44}$$

und kann gemittelt über alle Mikrophone angegeben werden als

$$\text{SNR}_{\text{Sensor}}(\Omega) = \frac{\frac{1}{M}\sum_{i=1}^{M}\phi_{S_iS_i}(\Omega)}{\frac{1}{M}\sum_{i=1}^{M}\phi_{N_iN_i}(\Omega)} = \frac{\text{Spur}\{\mathbf{\Phi_{SS}}(\Omega)\}}{\text{Spur}\{\mathbf{\Phi_{NN}}(\Omega)\}}, \tag{3.45}$$

wobei $\text{Spur}\{\boldsymbol{A}\}$ die Spur der Matrix $\boldsymbol{A}$ bezeichnet. Am *Beamformer*-Ausgang ergibt sich mit Gl. (3.21) folgender Ausdruck:

$$\text{SNR}_{\text{Array}}(\Omega) = \frac{\mathbf{F}^H(\Omega)\mathbf{\Phi_{SS}}(\Omega)\mathbf{F}(\Omega)}{\mathbf{F}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}(\Omega)}. \tag{3.46}$$

Mit Gl. (3.45) und Gl. (3.46) ergibt sich schließlich der SNR-Gewinn in Gl. (3.43) zu

$$G(\Omega) = \frac{\mathbf{F}^H(\Omega)\mathbf{\Phi_{SS}}(\Omega)\mathbf{F}(\Omega)}{\mathbf{F}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}(\Omega)} \cdot \frac{\text{Spur}\{\mathbf{\Phi_{NN}}(\Omega)\}}{\text{Spur}\{\mathbf{\Phi_{SS}}(\Omega)\}}. \tag{3.47}$$

Unter der Annahme, dass die unterschiedliche Dämpfung auf den Ausbreitungspfaden des Sprachsignals sowie Reflexionen vernachläßigt werden (Freifeldausbreitung), kann das Kreuzleistungsdichtespektrum vereinfacht werden zu

$$\mathbf{\Phi_{SS}}(\Omega)\Big|_{\substack{\mathbf{p}_t = \mathbf{p}_s \\ \mathbf{H}(\Omega) = \mathbf{d}(\Omega,\mathbf{p}_t)}} = \sigma_S^2(\Omega) \cdot \mathbf{d}(\Omega,\mathbf{p}_s)\mathbf{d}^H(\Omega,\mathbf{p}_s), \tag{3.48}$$

mit der Varianz des Sprachsignals $\sigma_S^2(\Omega)$. Der SNR-Gewinn kann dann für den Fall des unverzerrt gebliebenen Sprachsignals geschrieben werden als

$$G(\Omega)\Big|_{\substack{\mathbf{p}_t = \mathbf{p}_s \\ \mathbf{H}(\Omega) = \mathbf{d}(\Omega, \mathbf{p}_t)}} = \frac{|\mathbf{F}^H(\Omega)\mathbf{d}(\Omega, \mathbf{p}_s)|^2}{\mathbf{F}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}(\Omega)} \cdot \frac{\mathrm{Spur}\{\mathbf{\Phi_{NN}}(\Omega)\}}{M}. \tag{3.49}$$

Somit lassen sich bei gegebenen Filterkoeffizienten $\mathbf{F}(\Omega)$ Aussagen über die Störgeräuschreduktion für unterschiedliche Störschallfelder machen.

Der SNR-Gewinn innerhalb dieser Arbeit soll vorzugsweise im Zeitbereich ermittelt werden. Grundlage ist hierfür, dass in den Simulationen die einzelnen Komponenten der Eingangssignale, d. h. jeweils der Sprachanteil $s_i(n)$ und der Rauschanteil $n_i(n)$, separat vorliegen. So kann bei gegebenen Filterkoeffizienten der gefilterte Sprachanteil $y_s(n)$ und der gefilterte Rauschanteil $y_n(n)$ jeweils getrennt berechnet werden. Unter Beachtung der Menge der Zeitindizes T_s, welche Sprache beinhalten, soll folgende Definition gelten

$$\mathrm{SNRG} := 10 \cdot \left[\log_{10} \left(\frac{\sum\limits_{n \in T_s} y_s^2(n)}{\sum\limits_{n \in T_s} y_n^2(n)} \right) - \log_{10} \left(\frac{\sum\limits_{i=1}^{M} \sum\limits_{n \in T_s} s_i^2(n)}{\sum\limits_{i=1}^{M} \sum\limits_{n \in T_s} n_i^2(n)} \right) \right] \mathrm{dB}. \tag{3.50}$$

White Noise Gain Der so genannte *White Noise Gain* gibt den SNR-Gewinn für den Fall eines unkorrelierten Geräuschfeldes an. Da ein wesentlicher Grund für solch eine Störung Mikrophonrauschen sein kann (siehe Abbildung 3.1), ist dieser Wert also ein Gütemaß dafür, wie empfindlich der *Beamformer* auf Sensorrauschen reagiert. Für räumlich und zeitlich weißes Rauschen ergibt sich folgende Diagonalmatrix (diag$\{\cdot\}$) für das Kreuzleistungsdichtespektrum

$$\mathbf{\Phi}_{\mathbf{N}_u\mathbf{N}_u} = \mathrm{diag}\{\sigma^2_{\mathrm{N_{u,1}}}(\Omega), \sigma^2_{\mathrm{N_{u,2}}}(\Omega), \ldots, \sigma^2_{\mathrm{N_{u,M}}}(\Omega)\} \tag{3.51}$$

und unter der gerechtfertigten Annahme gleicher Varianzen $\sigma^2_{N_{u,1}}(\Omega) = \sigma^2_{N_{u,2}}(\Omega) = \ldots = \sigma^2_{N_{u,M}}(\Omega) = \sigma^2_{N_u}(\Omega)$ in den M Signalpfaden für das unkorrelierte Rauschen kann Gl. (3.51) weiter vereinfacht werden zu

$$\mathbf{\Phi}_{\mathbf{N}_u\mathbf{N}_u} = \sigma^2_{N_u}(\Omega) \cdot \mathbf{I}_M, \tag{3.52}$$

wobei mit $\mathbf{I}_M$ die Einheitsmatrix der Dimension M bezeichnet ist. Mit Gl. (3.52) kann der *White Noise Gain*

$$G^{\mathrm{W}}(\Omega) = G(\Omega)\Big|_{\mathrm{Weiß}} \tag{3.53}$$

angegeben werden zu

$$G^{\mathrm{W}}(\Omega)\Big|_{\substack{\mathbf{p}_t = \mathbf{p}_s \\ \mathbf{H}(\Omega) = \mathbf{d}(\Omega, \mathbf{p}_t)}} = \frac{|\mathbf{F}^H(\Omega)\mathbf{d}(\Omega, \mathbf{p}_s)|^2}{\mathbf{F}^H(\Omega)\mathbf{F}(\Omega)}. \tag{3.54}$$

und läßt sich für den *Uniformly Weighted Delay-and-Sum-Beamformer* weiter vereinfachen zu

$$G^{\mathrm{W}}_{\mathrm{DSB}}(\Omega)\Big|_{\substack{\mathbf{p}_t = \mathbf{p}_s \\ \mathbf{H}(\Omega) = \mathbf{d}(\Omega, \mathbf{p}_t)}} = \frac{|\mathbf{F}^H_{\mathrm{DSB}}(\Omega)\mathbf{d}(\Omega, \mathbf{p}_s)|^2}{\mathbf{F}^H_{\mathrm{DSB}}(\Omega)\mathbf{F}_{\mathrm{DSB}}(\Omega)} = M. \tag{3.55}$$

Für den *Uniformly Weighted Delay-and-Sum-Beamformer* ergibt sich also ein SNR-Gewinn für räumlich und zeitlich weißes Rauschen, das gleich der Anzahl der Mikrophone ist. Weiterhin bleibt festzuhalten, dass der *White Noise Gain* für alle anderen Filterkoeffizienten kleiner ausfällt, da bei gleichbleibender Norm von $\mathbf{F}(\Omega)$ der Ausdruck Gl. (3.54) und somit das innere Produkt $|\mathbf{F}^H(\Omega)\mathbf{d}(\Omega, \mathbf{p}_s)|$ maximal wird, wenn $\mathbf{F}(\Omega)$ und $\mathbf{d}(\Omega, \mathbf{p}_s)$ übereinstimmen.

Directivity Die Direktivität $D(\Omega)$ (engl. *Directivity*) gibt das Verhältnis der Leistung des aufgenommenen Schalls aus der *Array*-Blickrichtung im Verhältnis zur Schallleistung aus allen Raumrichtungen[9] (θ, φ) abhängig von der Frequenz an:

$$D(\Omega) = \frac{|B(\Omega, \theta_t, \varphi_t)|^2}{\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} |B(\Omega, \theta, \varphi)|^2 \sin\theta d\theta d\varphi}. \tag{3.56}$$

Die formale Darstellung der Direktivität in Gl. (3.56) kann nach Berechnung der Integrale (vgl. Anhang B) als Funktion der Kohärenzmatrix $\mathbf{\Gamma}(\Omega)$ eines diffusen Schallfelds dargestellt werden

$$D(\Omega) = \frac{|\mathbf{F}^H(\Omega)\mathbf{d}(\Omega, \mathbf{p}_t)|^2}{\mathbf{F}^H(\Omega)\mathbf{\Gamma}(\Omega)\mathbf{F}(\Omega)}, \tag{3.57}$$

wobei die Matrixelemente $\Gamma_{i,j}(\Omega)$ nach Gl. (2.20) zu besetzen sind. Die Direktivität in Gl. (3.57) kann derart interpretiert werden, dass sie dem SNR-Gewinn Gl. (3.49) des *Arrays* im Falle eines diffusen Störschallfeldes

$$G(\Omega)\Big|_{\text{Diffus}} = G^D(\Omega) \tag{3.58}$$

und einem empfangenen Sprachsignals ohne Hallkomponenten entspricht:

$$D(\Omega) = G^D(\Omega)\Bigg|_{\substack{\mathbf{p}_t = \mathbf{p}_s \\ \mathbf{H}(\Omega) = \mathbf{d}(\Omega, \mathbf{p}_t)}} . \tag{3.59}$$

Das Bündelungsmaß (engl. *Directivity Index*) gibt die zur Direktivität äquivalente Darstellung im logarithmischen Maß an:

$$DI(\Omega) := 10\log_{10}\big(D(\Omega)\big)\text{dB}. \tag{3.60}$$

Für die einfachste Wahl der Filterkoeffizienten als DSB ($\mathbf{F}_{\text{DSB}}(\Omega)$) ergeben sich für das Bündelungsmaß die in Bild 3.6 und 3.7 gezeigten Verläufe. Dabei ist das Bündelungsmaß jeweils über der Frequenz für verschiedene Zielrichtungen θ_t in Bild 3.6 und für unterschiedliche Mikrophonanzahl/-abstands-Kombinationen in Bild 3.7 aufgetragen. Es ergibt sich dabei ein wellenförmiger Verlauf des *Directivity Index*, der mit dem si-förmigen Verlauf der Kohärenz korrespondiert, d. h. der $DI(\Omega)$ schwingt um $10\log(M)$ herum. Der $DI(\Omega)$ steigt mit größer werdendem Mikrophonabstand steiler an; es wird also ein höherer SNR-Gewinn bei niedrigen Frequenzen erzielt. Weiterhin nimmt der $DI(\Omega)$ mit zunehmender Mikrophonanzahl zu.

[9] An dieser Stelle wird das in Gl. (3.36) eingeführte *Beampattern* in Abhängigkeit vom Elevationswinkel φ und vom Azimuthwinkel θ für planar einfallende Schallwellen geschrieben.

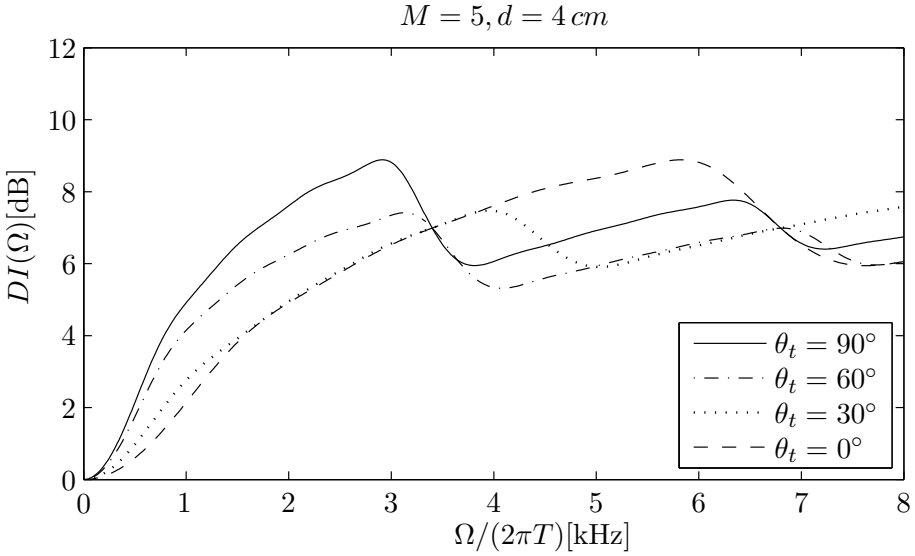

Bild 3.6: Unterschiedliche Verläufe für das Bündelungsmaß abhängig von der Zielrichtung θ_t aufgetragen über der Frequenz für $M = 5$ Mikrophone mit äquidistantem Abstand von $d = 4\,\text{cm}$.

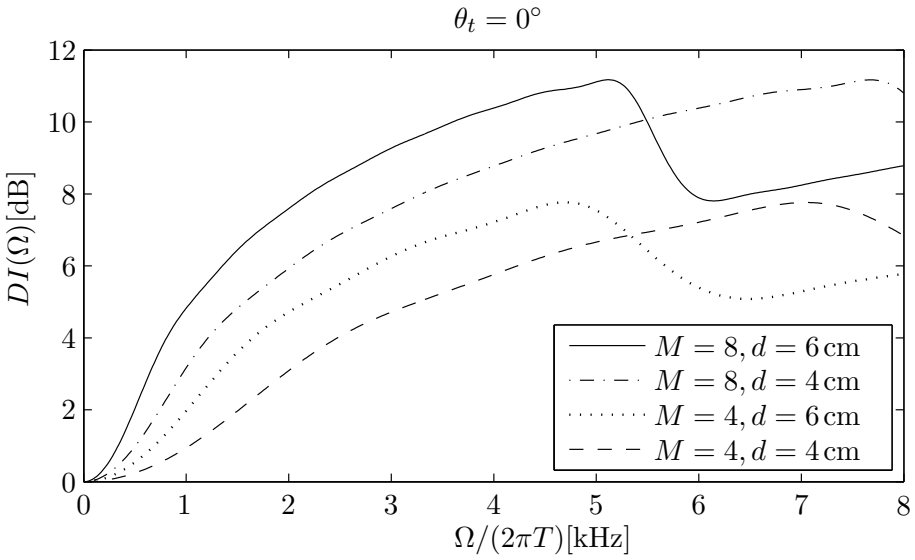

Bild 3.7: Unterschiedliche Verläufe für das Bündelungsmaß bei einer Zielrichtung $\theta_t = 0°$ aufgetragen über der Frequenz für unterschiedliche Kombinationen zwischen der Mikrophonanzahl $M \in \{4, 8\}$ und dem Mikrophonabstand $d \in \{4\,\text{cm}, 6\,\text{cm}\}$.

Averaged SNR Ein wichtiges Hilfsmittel zur Schätzung des Signal-zu-Rauschabstandes ist das gemittelte SNR (engl. *averaged* SNR) im Falle stationärer Störgeräusche [GBW01]. Dabei liegt das Sprachsignal nicht separat in reiner Form vor, sondern muss mit Hilfe des Mischsignals bestehend aus Sprach-plus-Rauschanteil geschätzt werden [WHU06b]. Dazu ist es notwendig, mit Hilfe einer Sprache/Pause-Detektion (engl. *Speech Activity Detection*, VAD) Signalsegmente mit bzw. ohne Sprachanteil entsprechend zu klassifizieren. So ergibt sich z. B.

für das gemittelte SNR am *Beamformer*-Ausgang folgende Beziehung

$$\mathrm{SNR}_{\mathrm{avg}} = \frac{\frac{1}{L_s}\sum\limits_{n\in T_s} y^2(n) - \frac{1}{L_n}\sum\limits_{n\in T_n} y^2(n)}{\frac{1}{L_n}\sum\limits_{n\in T_n} y^2(n)}. \tag{3.61}$$

Mit L_s ist die Anzahl der Abtastwerte bezeichnet, die zusätzlich zum Rauschen auch das Sprachsignal enthalten, und mit L_n die Anzahl der Abtastwerte, in denen lediglich Rauschen beobachtet wird. Weiter bezeichnet T_s die Menge der Zeitindizes, welche Sprache, und T_n die Menge der Zeitindizes, welche keine Sprache beinhalten.

3.5 Wahrnehmungsbasierte Qualitätsbewertung des Sprachsignals

Eine aussagekräftige subjektive Beurteilung der Sprachqualität verarbeiteter Signale läßt sich durch Auswertung von Hörtests einer Gruppe von Versuchspersonen angeben. Die Internationale Fernmeldeunion (International Telecommunication Union, ITU) hat dafür Bewertungsmethoden spezifiziert, welche unter dem Begriff "*Mean Opinion Score*"[10] (MOS) zusammengefasst sind. Da solche Hörtests mit hohem Aufwand verbunden sind, werden häufig objektive Beurteilungsverfahren herangezogen, um auf der Basis von Algorithmen eine quantitative Aussage über die Qualität der verarbeiteten Signale angeben zu können. Häufig verwendete Methoden lassen sich dabei prinzipiell in zwei Klassen unterteilen. Zum einen sind dies Maße basierend auf dem Vergleich von Sprachmodellparametern, die mit Hilfe der Methode der linearen Prädiktion gewonnen werden, wie z. B. *Itakura-Saito-Distortion*, *Log-Likelihood-Ratio* oder *Log-Area-Ratio*, siehe [IS70, GM76, QBC88]. Und zum anderen Verfahren, welche Modelle der auditorischen Signalverarbeitung nutzen, wie z. B. *Perceptual Evaluation of Speech Quality* (PESQ) [ITU01] und das Perzeptive Modell zur Qualitätsbeurteilung (PEMO-Q) [Hub03]. Im Vergleich zu den erstgenannten objektiven Bewertungsverfahren hat sich die PEMO-Q-Methode als sehr gute Alternative erwiesen [RHK05, Hub06]. Daher wird in dieser Arbeit das PEMO-Q-Verfahren verwendet, um wahrnehmungsbasierte Unterschiede zwischen Audiosignalen anzugeben.

Bei der instrumentellen Methode PEMO-Q besteht die Grundidee darin, basierend auf dem Gehörmodell der "effektiven" auditorischen Signalverarbeitung nach [DPK96], die zu vergleichenden akustischen Signale in interne Repräsentationen auf perzeptueller Ebene zu überführen. Die Korrelation der internen Repräsentationen beider Signale ist dann ein Maß für die wahrgenommene Ähnlichkeit dieser Signale: *Perceptual Similarity Measure* (PSM). Jeder wahrnehmbare Unterschied wird als Qualitätsverschlechterung des Testsignals gegenüber dem Referenzsignal interpretiert.

Das PEMO-Q-Verfahren zum Vergleich eines Test- und Referenzsignales läßt sich prinzipiell in 4 Verarbeitungsstufen einteilen (siehe Bild 3.8):

1.) Vorverarbeitung:
Vor der Transformation der Signale in interne Repräsentationen kann eine zeitliche

[10] Der MOS bezeichnet Verfahren zur subjektiven Beurteilung der Qualität von Sprach- und Bildübertragungen, welche in der ITU-Empfehlung P.800 spezifiziert sind und in der Empfehlung P.830 werden die Bewertungsmethoden aktuell verfeinert.

Verschiebung sowie eine Pegeldifferenz zwischen den Signalen ermittelt und ausgeglichen werden. Weiterhin können Pause-Segmente herausgeschnitten und somit aus der Messung herausgehalten werden.

2.) Transformation in neuronale Aktivitätsmuster:
Psychoakustisch motiviert erfolgt zunächst eine Aufteilung in 33 Bänder mittels einer Gammaton-Filterbank entsprechend der Basilarmembran-Bandpasscharakteristik mit Mittenfrequenzen zwischen 235 Hz und 14,5 kHz. Danach werden die Frequenzbänder unabhängig voneinander weiterverarbeitet; zuerst durch eine Halbwellen-Gleichrichtung und eine 1 kHz Tiefpassfilterung, welche die Transformation der mechanischen Oszillation der Schallwellen in neuronales Feuern der inneren Hörzellen simuliert. Anschließend werden psychoakustische Effekte bezüglich zeitlicher Maskierung und Adaption durch fünf aufeinander folgende mittels Division rückgekoppelte Tiefpassfilter modelliert. Dadurch werden sich schnell ändernde Signale stärker hervorgehoben im Vergleich zu stationären Signalanteilen.

3.) Nachverarbeitung:
Die Einhüllende wird mittels einer 8-kanaligen linearen Modulationsfilterbank ermittelt, so dass schließlich die $33 \cdot 8 = 264$ Ausgänge die so genannte interne Repräsentation des akustischen Signals bilden. Im Falle betragsmäßig kleinerer Repräsentanten für das Testsignal im Vergleich zum Referenzsignal wird der interne Repräsentant des Testsignals durch Mittelung beider ersetzt. Dieser Verarbeitungsschritt ist motiviert durch die Annahme, dass fehlende Komponenten im Signal weniger störend wirken als zusätzlich eingefügte Geräuschartefakte.

4.) Korrelation:
Die über die Zeit und Frequenz gemittelten Kreuzkorrelationen zwischen jedem Repräsentanten des Test- und Referenzsignals werden auf das Intervall [-1,1] normiert und ergeben schließlich den PSM-Wert.

Weiterhin ist es mit PEMO-Q möglich, die interne Repräsentation auf eine 5-stufige wahrnehmungsbasierte Skala zu transformieren und die Differenz als *Objective Difference Grade* (ODG) anzugeben[11]. Dabei ist die Beeinträchtigung der Audioqualität entsprechend der ITU-Empfehlung[12] eingeteilt.

Bevor das Pemo-Q-Verfahren in späteren Kapiteln zur Sprachqualitätsbeurteilung benutzt wird, soll im Folgenden beispielhaft einerseits die Auswirkung einer fehlerhaften Laufzeitkompensation auf das Sprachsignal am Ausgang eines DSBs (unter der Annahme einer Schallausbreitung im Freifeld) und andererseits der Einfluss von Nachhall auf ein unverzerrtes Sprachsignal untersucht werden.

Sprachverzerrung durch fehlerhafte Laufzeitkompensation

In einem DSB können zwei unterschiedliche Fehlerquellen dazu führen, dass die Sprachkomponenten in den Mikrophonpfaden nicht exakt kohärent aufaddiert werden. Einerseits ergibt sich offensichtlich eine fehlerhafte Laufzeitkompensation durch einen Lokalisationsfehler des

[11] Das PEMO-Q-Softwarepaket liefert noch die weiteren Qualitätsmaße *Qc* nach [HK00], den instantanen PSM-Wert PSM(t) und einen lautheitsgewichteten Verlauf der instantanen PSM-Werte.

[12] Subjektives Qualitätsmaß nach den ITU-Empfehlungen BS.562-3: *Subjective Assessment of Sound Quality.*

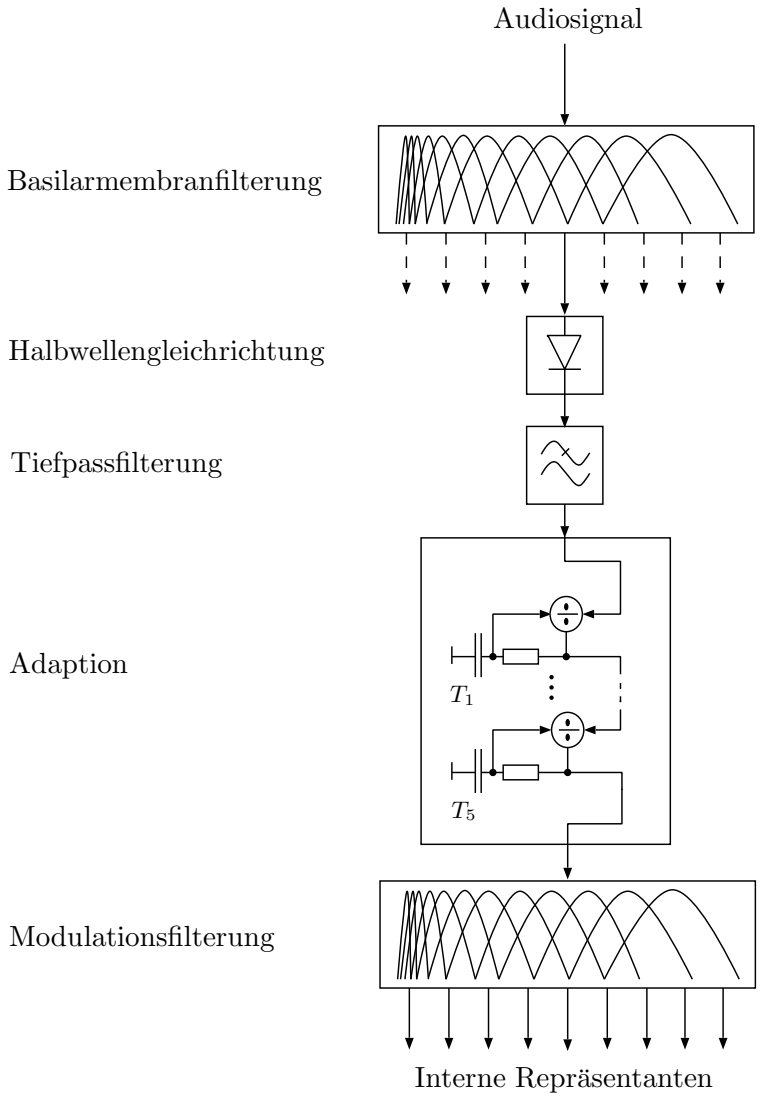

Bild 3.8: Blockschaltbild des auditorischen Modells zur Qualitätsbeurteilung.

Sprechers. Andererseits resultiert aber auch bei korrekt ermittelter Sprecherrichtung ein ungenau eingestellter Mikrophonabstand in falsch berechneten Kompensationszeiten. Betrachtet man zur Anschauung die in den Bildern 3.3 und 3.4 dargestellten Richtcharakteristiken unter dem Gesichtspunkt eines eventuell aufgetretenen Lokalisationsfehlers, so wird klar, das lineare Verzerrungen des Sprachsignals durch eine frequenzabhängige Dämpfung des Quellsignals auftreten. Diese Dämpfung wächst grundsätzlich mit steigender Frequenz aufgrund der schmaler werdenden Hauptkeule an. Analytisch kann die Sprachsignaldämfung durch Auswertung des *Beampatterns* in Gl. (3.41) für die tatsächliche Sprecherrichtung $\theta = \theta_s$ erfolgen, wobei die Sprecherrichtung mit der Ausrichtung des *Arrays* über die Abweichung $\Delta\theta$ zusammenhängen soll:

$$\theta_t = \theta_s + \Delta\theta. \tag{3.62}$$

Dazu soll die effektive Verzögerung in Gl. (3.40) ausgeschrieben werden zu

$$\begin{aligned}\tau_e &= \frac{1}{c}\left(d\sin(\theta_t) - d\sin(\theta)\right) && (3.63)\\ &= \frac{1}{c}\left(d\sin(\theta_s+\Delta\theta) - d\sin(\theta_s)\right) && (3.64)\\ &\stackrel{!}{=} \frac{1}{c}\left((d+\Delta d(\theta_s,\Delta\theta))\sin(\theta_s) - d\sin(\theta_s)\right), \qquad \theta_s \neq 0. && (3.65)\end{aligned}$$

Durch das Gleichsetzen von Gl. (3.64) mit Gl. (3.65) soll angedeutet sein, dass eine fehlerhafte Lokalisation zu der gleichen effektiven Verzögerung führt wie eine fehlerhafte Anordnung der Mikrophone. Die Mikrophone befinden sich also in dem tatsächlichen Abstand von $d + \Delta d(\theta_s, \Delta\theta)$ anstatt des angenommenen Abstandes d zueinander. Ein zu einem Lokalisationsfehler äquivalentes $\Delta d(\theta_s, \Delta\theta)$ kann allerdings nur für eine Sprecherrichtung $\theta_s \neq 0$ angegeben werden, da für eine *Broadside*-Ausrichtung keine Signalverzögerung notwendig ist und für beliebige Mikrophonabstände[13] die Summation der Mikrophonsignale das korrekte Ergebnis liefert. In Bild 3.9 ist die Sprachsignaldämpfung durch Auswertung von Gl. (3.41) mit den Annahmen Gl. (3.62) bis Gl. (3.65) dargestellt:

$$B_{\mathrm{DSB}}(\Omega, \theta_s; \theta_t = \theta_s + \Delta\theta) = B^{\mathrm{LE}}_{\mathrm{DSB}}(\Omega, \Delta\theta). \tag{3.66}$$

Mit dem Index "LE" soll hierbei die Auswertung des *Beampatterns* bezüglich eines Lokalisationsfehlers (engl. *Localization Error*, LE) angedeutet sein. In Bild 3.9 (a) ist die Sprachsignaldämpfung für eine Zielausrichtung $\theta_t = 0°$ und in Bild 3.9 (b) für $\theta_t = 60°$ dargestellt. In Bild 3.10 ist der äquivalente fehlerhafte Abstand $\Delta d(\theta_s, \Delta\theta)$ über dem korrespondierenden Lokalisationsfehler $\Delta\theta$ für unterschiedliche Sprecherrichtungen aufgetragen. Die in Bild 3.9

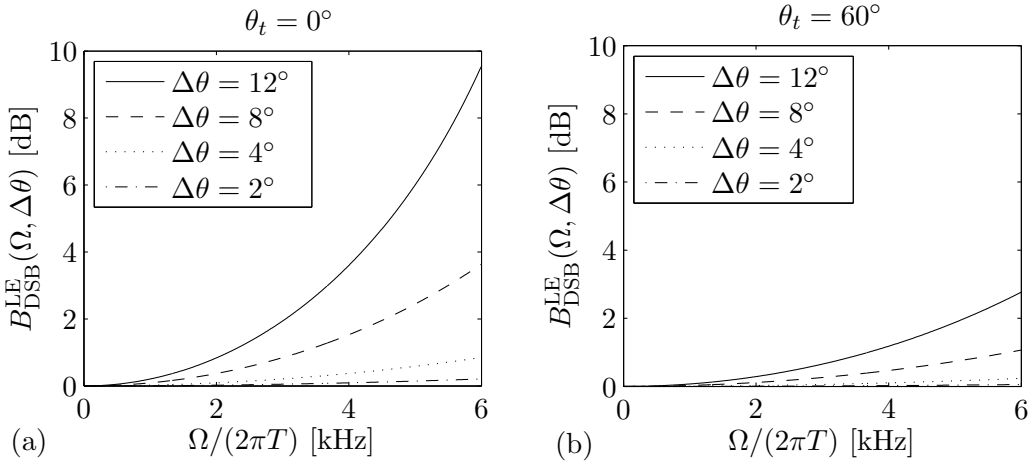

Bild 3.9: Unterschiedliche Verläufe der Sprachsignaldämpfung in Abhängigkeit von der Frequenz für einen DSB. In (a) für eine Zielausrichtung $\theta_t = 0°$ und in (b) für $\theta_t = 60°$ jeweils für $M = 5$ Mikrophone mit äquidistantem Abstand von $d = 4\,\mathrm{cm}$.

gezeigten Verläufe der Sprachsignaldämpfung zeigen deutlich ein frequenzselektives Verhalten. Dies ist offensichtlich, da, wie in Bild 3.4 bereits gezeigt wurde, die Breite der Hauptkeule zu höheren Frequenzen hin immer schmaler wird und sich so ein Lokalisationsfehler dort besonders stark auswirkt. Weiterhin ist der Effekt der Sprachsignaldämpfung bei gleichem

[13] Die Aussage, dass falsch angenommene Mikrophonabstände bei einem von *Broadside*-Richtung einfallenden Sprachsignal keinerlei Auswirkung auf das resultierende Sprachsignal hat, gilt natürlich nur, solange die Fernfeld-Annahme Gültigkeit hat.

Lokalisationsfehler für verschiedene Zielrichtungen θ_t unterschiedlich stark ausgeprägt. Auch dieses Verhalten kann durch einen Vergleich mit Bild 3.4 erklärt werden: Die Breite der Hauptkeule nimmt bei gleicher Frequenz für Zielrichtungen von einer *Broadside*-Ausrichtung hin zur *Endfire*-Ausrichtung weiter zu, wodurch sich Lokalisationsfehler in einer geringer werdenden Dämpfung des Sprachsignals bemerkbar machen.

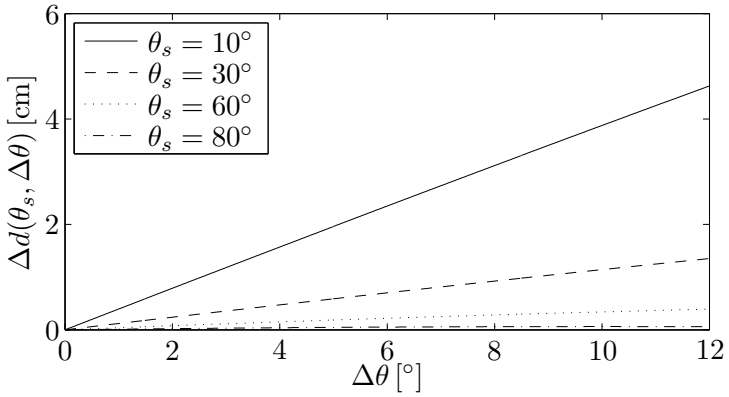

Bild 3.10: Äquivalenter Fehlerabstand $\Delta d(\theta_s, \Delta\theta)$ in Abhängigkeit vom Lokalisationsfehler $\Delta\theta$ für unterschiedliche Sprecherrichtungen; jeweils für $M = 5$ Mikrophone mit äquidistantem Abstand von $d = 4\,\text{cm}$.

Für den äquivalenten Fehlerabstand $\Delta d(\theta_s, \Delta\theta)$ kann bezüglich der Ausrichtung gefolgert werden, dass der negative Effekt einer Sprachsignaldämpfung sich um so stärker auswirkt, je weiter die Sprecherrichtung von *Broadside* hin zu *Endfire* übergeht (bei gleichbleibendem $\Delta d(\theta_s, \Delta\theta)$).

Ebenfalls destruktiv auf das Sprachsignal am *Beamformer*-Ausgang wirkt sich eine unterschiedliche Dämpfung des Sprachsignals durch unterschiedliche Ausbreitungspfade zu den einzelnen Mikrophonen (falls diese nicht entsprechend kompensiert wird) auf das DSB-Ausgangssignal aus. Ein ähnlicher Effekt stellt sich ein, wenn ein signifikanter Unterschied zwischen den Mikrophoncharakteristiken vorliegt und dadurch ein systematischer Fehler in der Pegelgewichtung entsteht [DM99]. Dieser Effekt fällt jedoch weitaus geringer als ein Lokalisationsfehler aus und wird daher hier nicht weiter untersucht.

Nach den bisherigen Betrachtungen zur frequenzselektiven Signaldämpfung scheint die Auswirkung eines Lokalisationsfehlers auf das Sprachsignal erheblich zu sein. Da jedoch die spektrale Leistungsdichte von Sprachanteilen im oberen Frequenzbereich gering im Vergleich zu den stimmhaften Anteilen im unteren Frequenzbereich ist, fällt eine fehlerhafte Ausrichtung bei einer subjektiven Bewertung der Qualität des Sprachsignals deutlich geringer ins Gewicht, als dies durch die Verläufe in Bild 3.9 vermutet wird. Diese Wahrnehmung spiegelt sich ebenfalls in der Qualitätsbeurteilung nach dem PEMO-Q-Verfahren wieder. In Bild 3.11 (a) ist beispielhaft der Verlauf der PSM-Werte in Abhängigkeit von dem Lokalisationsfehler $\Delta\theta$ dargestellt. Dabei wurden für 10 Sprachbeispiele (5 männliche und 5 weibliche Sprecher, abgetastet mit einer Frequenz von 12 kHz) $M = 5$-kanalige Signale unter der Annahme einer Schallausbreitung im Freifeld jeweils für unterschiedliche Einfallsrichtungen auf die Sensorgruppe simuliert. Diese wurden mittels DSB mit *Broadside*-Ausrichtung verarbeitet und die einkanaligen Ausgangsdaten im Vergleich zu den Referenzsignalen bei $\Delta\theta = 0\,°$ bezüglich der perzeptiven Sprachqualität verglichen. Das Bild 3.11 (a) zeigt die PSM-Ergebnisse jeweils gemittelt über die 10 verwendeten Sprachbeispiele. In dem Bild 3.11 (b) ist die spektrale Leistungsdichte des Ausgangssignals für unterschiedliche Lokalisationsfehler über der Fre-

quenz aufgetragen; ebenfalls gemittelt über alle Sprachbeispiele. Der Vergleich von Bild 3.11 mit Bild 3.9 zeigt zwar, dass die relative, frequenzselektive Sprachsignaldämpfung dem theoretischen Verlauf entspricht, aber der messbare Qualitätsverlust der Sprache aufgrund der niedrigen Leistung in den höheren Frequenzen sehr gering ist.

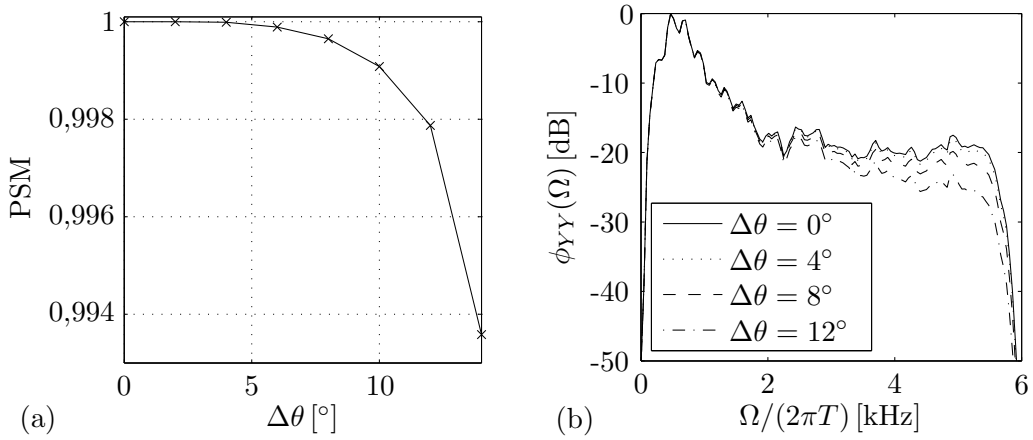

Bild 3.11: In (a) wahrnehmungsbasierte Qualitätsbeurteilung und in (b) spektrale Leistungsdichte, jeweils für die Ausgangssignale eines DSBs für unterschiedliche Lokalisationsfehler $\Delta\theta$ mit *Broadside*-Ausrichtung von $M = 5$ Mikrophonen bei äquidistantem Abstand $d = 4\,\mathrm{cm}$.

Einfluss von Hall auf PEMO-Q

Abschließend soll nun noch einerseits der negative Einfluss von Hall und andererseits die positive, enthallende Wirkung der kohärenten Überlagerung mehrkanaliger Sprachsignale[14] bezüglich der wahrnehmungsbasierten Qualitätsbewertung mittels des PEMO-Q-Verfahrens gezeigt werden. In Bild 3.12 ist beispielhaft der Verlauf der PSM-Werte in Abhängigkeit von der Nachhallzeit T_{60} dargestellt. Verglichen werden hierbei die unverhallten 10 Sprachbeispiele mit den jeweils verhallten Versionen dieser Referenzsignale. Dabei wurde mit der Spiegelquellenmethode in einem Raum der Größe $(6\,\mathrm{m})\,\mathrm{x}\,(5\,\mathrm{m})\,\mathrm{x}\,(3\,\mathrm{m})$ für unterschiedliche Nachhallzeiten zwischen 0 s und 0,8 s jeweils die Schallausbreitung zwischen einer Sprachsignalquelle und fünf Sensoren mit dem Abstand von 0,8 m zum *Array*-Mittelpunkt simuliert. Das Bild 3.12 zeigt die PSM-Ergebnisse jeweils gemittelt über die 10 verwendeten Sprachbeispiele für das mittlere der fünf Sensoren (DSB-Eingangssignale) gekennzeichnet durch "Mik" und den DSB-Ausgang "DSB". Die Sprecherrichtung ist dabei gleich der *Beamformer*-Ausrichtung $\theta_s = \theta_t = 0°$.

Anhand der starken Auswirkung von Hall auf die gemessenen PSM-Werte (vgl. Bild 3.11 mit 3.12) erscheint es sinnvoll, in späteren Vergleichen zur Sprachverzerrung jeweils Referenzsignale heranzuziehen, welche sehr ähnliche Halleigenschaften wie die zu testenden Signale aufweisen. Dafür werden dann jeweils mittels eines Referenzsystems optimal gefilterte verhallte Sprachsignale als Referenzsignale für die zu vergleichenden *Beamforming*-Verfahren genutzt.

[14] Bei kohärenter Überlagerung mehrkanaliger verhallter akustischer Signale steigt der Energieanteil der Schallausbreitung über die direkte Komponente der resultierenden Raumimpulsantwort und somit das Klarheitsmaß des Sprachsignals, vgl. Abschnitt 2.2.

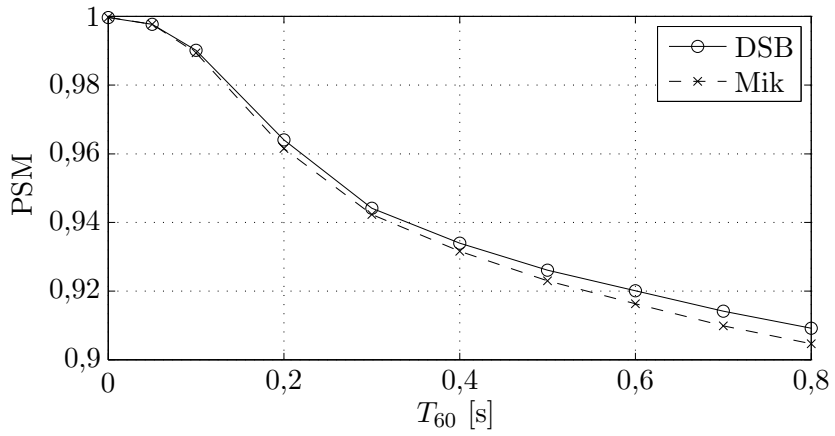

Bild 3.12: Degradation der perzeptuellen Qualitätsbeurteilung von verhallten Sprachsignalen "Mik" mit den jeweils unverhallten Versionen als Referenz im Vergleich zur enthallenden Wirkung eines DSBs "DSB".

3.6 Zusammenfassung

In diesem Kapitel wurden die grundlegenden Eigenschaften von Mikrophongruppen und deren Bewertungskriterien aufgezeigt. Hierbei kamen ausschließlich Betrachtungen von linear angeordneten *Arrays* zum Einsatz, wie sie auch im weiteren Verlauf dieser Arbeit als Ausgangspunkt für das anschließende *Beamforming* dienen sollen. Mit dem eingeführten *Beamformer*-Signalmodell wurde anhand eines laufzeitkompensierenden so genannten *Delay-and-Sum-Beamformers* die Auswirkung räumlichen *Aliasings* in Abhängigkeit von dem gewählten Mikrophonabstand untersucht. Dabei ergibt sich einerseits ein bestimmter maximaler Mikrophonabstand, um räumliches *Aliasing* zu vermeiden. Andererseits ist ein deutlich größerer Abstand wünschenswert, um tiefe Frequenzen des Störschallfeldes stärker unterdrücken zu können. Folglich soll als Kompromiss zwischen den beiden gegensätzlichen Kriterien im überwiegenden Teil der Arbeit ein Mikrophonabstand von $d = 4\,\text{cm}$ bei einer Abtastrate von $f_{Ab} = 12\,\text{kHz}$ zum Einsatz kommen.

Zur Analyse des räumlichen *Aliasings* wurde die Richtcharakteristik der Raum-Zeit-Filterung mittels *Beamforming*-Verfahren eingeführt, welche eine räumliche Übertragungsfunktion für Schallwellen aus den entsprechend zu analysierenden Raumrichtungen darstellt. Die Richtcharakteristik (bzw. *Beampattern*) ist ein wichtiges Werkzeug zur Veranschaulichung und zur Leistungsbeurteilung von *Beamformern* bezüglich ihrer räumlichen Selektivität.

Als weitere Bewertungsgrößen der erzielbaren Geräuschreduktion mittels *Beamforming* wurden einerseits wahrnehmungsbasierte Qualitätsmerkmale und andererseits SNR-basierte Bewertungskriterien beschrieben. Hierbei kann die SNR-Verbesserung vom Ein- zum Ausgang des *Arrays* (auch *Array Gain* oder SNR *Gain*) unterschieden werden für den Fall von weißem, räumlich unkorrelierten Rauschen (*White Noise Gain*) und diffusem Rauschen (*Directivity*) als Störschallfeld.

Für die wahrnehmungsbasierte Qualitätsbewertung mittels des PEMO-Q-Verfahrens über den PSM-Wert wurden in diesem Kapitel erste Ergebnisse für den *Delay-and-Sum-Beamformer* zum einen für die Annahme von Lokalisationsfehlern durchgeführt, und zum anderen Analysen zur Abschätzung der Hall-Auswirkung mit und ohne *Beamformer* vorgenommen. Dabei zeigte sich, dass die Auswirkung von Lokalisationsfehlern sowohl bei subjektiven Hörtests als

auch bei Verwendung des Ähnlichkeitsmaß geringer ausfallen als vermutet. Verhallte Signale zeigten hingegen eine hohe Abweichung in der PSM-Bewertung bezüglich einer unverhallten Referenz. Die Verarbeitung eines verhallten mehrkanaligen Sprachsignals mittels DSB zeigte hier wie erwartet eine messbare Verbesserung (Enthallung) des Signals.

Kapitel 4

Statistisch optimales Beamforming

Im Gegensatz zu dem bisher betrachteten *Delay-and-Sum-Beamformer* werden im folgenden Kapitel die Grundlagen für das so genannte statistisch optimale *Beamforming* hergeleitet. Dabei erfolgt die Wahl der *Beamformer*-Gewichtungsvektoren basierend auf den statistischen Eigenschaften des Sprachsignals und des Störschallfelds. Zunächst soll hier davon ausgegangen werden, dass die Eingangssignale zumindest schwach stationär sind und deren Statistik zweiter Ordnung bekannt ist. Auf den praktisch relevanten Fall unbekannter Signalstatistik bzw. sich zeitlich ändernder Signaleigenschaften wird in diesem Kapitel nur peripher eingegangen. Diese Problematik ist vielmehr Gegenstand der weiteren Kapitel, in denen es um die adaptive Berechnung der Filtergewichte geht.

Zunächst sollen die Filterkoeffizienten derart bestimmt werden, so dass das frequenzabhängige Schmalband-SNR maximiert wird. Dieses so genannte Max-SNR-Kriterium führt zu einem verallgemeinerten Eigenwertproblem, wobei die optimalen Filterkoeffizienten gerade durch den Eigenvektor korrespondierend zum größten Eigenwert des vorliegenden Eigenwertproblems gegeben sind. Es soll gezeigt werden, dass eine Skalierung der resultierenden Filterkoeffizienten durch eine einkanalige Nachfilterung (engl. *Post Filter*) identisch zu Lösungen ist, welche über andere Optimierungskriterien hergeleitet werden können. Diese Kriterien sind insbesondere Minimierung der Varianz (engl. *Minimum Variance*, MV), Maximierung der Plausibilität (engl. *Maximum Likelihood*, ML) und Minimierung des kleinsten mittleren quadratischen Fehlers (engl. *Minimum Mean Squared Error*, MMSE).

4.1 Max-SNR

Es sollen nun die optimalen Filterkoeffizienten derart hergeleitet werden, so dass das frequenzabhängige SNR am Ausgang des *Arrays*

$$\mathrm{SNR}_{\mathrm{Array}}(\Omega) = \frac{\mathbf{F}^H(\Omega)\mathbf{\Phi_{SS}}(\Omega)\mathbf{F}(\Omega)}{\mathbf{F}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}(\Omega)} \tag{4.1}$$

maximiert wird. Offensichtlich stellt der Quotient in Gl. (4.1) den so genannten Rayleigh Quotienten bezüglich der Matrizen $\mathbf{\Phi_{SS}}(\Omega)$ und $\mathbf{\Phi_{NN}}(\Omega)$ dar [Hay02]. Bei den betrachteten Matrizen der Kreuzleistungsdichtespektren (KLDS) handelt es sich in der Regel um positiv definite Matrizen[1]. Daher kann gefolgert werden, dass die Eigenwerte des verallgemeinerten

[1] Im praktischen Fall der messtechnischen, iterativen Bestimmung der KLDS-Matrizen handelt es sich aufgrund unkorrelierter Rauschterme in den Signalpfaden um positiv definite Matrizen. Dennoch kann hier zur

Eigenwertproblems (engl. *Generalized Eigenvalue Problem*, GEVP) positiv und reellwertig sind, und dass sich das SNR Gl. (4.1) in dem Bereich

$$0 < \mathrm{SNR}_{\mathrm{Array}}(\Omega) \leq \lambda_S^{(\max)}(\Omega). \tag{4.2}$$

bewegt. In Gl. (4.2) ist mit $\lambda_S^{(\max)}(\Omega)$ der größte frequenzabhängige Eigenwert bezeichnet, der zum verallgemeinerten Eigenwertproblem gehört. Dieser Wert wird genau dann erreicht, wenn der Koeffizientenvektor $\mathbf{F}(\Omega)$ gerade so gewählt wird, dass er einem Eigenvektor $\mathbf{F}^{(\max)}(\Omega)$ korrespondierend zum größten Eigenwert $\lambda_S^{(\max)}(\Omega)$ entspricht; dann wird das SNR maximiert zu

$$\mathrm{SNR}_{\mathrm{Array}}^{(\max)}(\Omega) = \frac{{\mathbf{F}^{(\max)}}^H(\Omega)\mathbf{\Phi_{SS}}(\Omega)\mathbf{F}^{(\max)}(\Omega)}{{\mathbf{F}^{(\max)}}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}^{(\max)}(\Omega)} = \lambda_S^{(\max)}(\Omega). \tag{4.3}$$

Unter Verwendung der verallgemeinerten Eigenwertgleichung

$$\begin{aligned} \mathbf{\Phi_{SS}}(\Omega)\mathbf{F}^{(\max)}(\Omega) &= \lambda_S^{(\max)}(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}^{(\max)}(\Omega) && (4.4)\\ &= \frac{{\mathbf{F}^{(\max)}}^H(\Omega)\mathbf{\Phi_{SS}}(\Omega)\mathbf{F}^{(\max)}(\Omega)}{{\mathbf{F}^{(\max)}}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}^{(\max)}(\Omega)}\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}^{(\max)}(\Omega) && (4.5)\end{aligned}$$

und der Annahme räumlich und zeitlich stationärer Sprachsignale

$$\mathbf{\Phi_{SS}}(\Omega) = \phi_{S_cS_c}(\Omega)\mathbf{H}(\Omega)\mathbf{H}^H(\Omega) \tag{4.6}$$

lässt sich der optimale Koeffizientenvektor $\mathbf{F}^{(\max)}(\Omega)$ analytisch berechnen:

$$\mathbf{H}(\Omega) = \frac{{\mathbf{F}^{(\max)}}^H(\Omega)\mathbf{H}(\Omega)}{{\mathbf{F}^{(\max)}}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}^{(\max)}(\Omega)}\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}^{(\max)}(\Omega). \tag{4.7}$$

Für Gl. (4.7) ist ausgenutzt worden, dass der Skalar $\phi_{S_cS_c}(\Omega)\mathbf{H}^H(\Omega)\mathbf{F}^{(\max)}(\Omega)$ auf beiden Seiten der Gleichung Gl. (4.5) nach Einsetzen von Gl. (4.6) vorhanden ist und daher gekürzt werden kann. Es folgt weiter

$$\begin{aligned} \mathbf{F}^{(\max)}(\Omega) &= \frac{{\mathbf{F}^{(\max)}}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}^{(\max)}(\Omega)}{{\mathbf{F}^{(\max)}}^H(\Omega)\mathbf{H}(\Omega)}\mathbf{\Phi_{NN}}^{-1}(\Omega)\mathbf{H}(\Omega) && (4.8)\\ &= \frac{{\mathbf{F}^{(\max)}}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}^{(\max)}(\Omega)}{{\mathbf{F}^{(\max)}}^H(\Omega)\mathbf{H}(\Omega)}\mathbf{F}_{\mathrm{SNR}}(\Omega), && (4.9)\end{aligned}$$

wobei folgende Definition gelten soll

$$\mathbf{F}_{\mathrm{SNR}}(\Omega) := \mathbf{\Phi_{NN}}^{-1}(\Omega)\mathbf{H}(\Omega). \tag{4.10}$$

Betrachtet man Gl. (4.9) und Gl. (4.3), so ist festzustellen, dass der skalare Faktor vor dem Vektor $\mathbf{F}_{\mathrm{SNR}}(\Omega)$ in Gl. (4.9) bezüglich des SNRs keine Rolle spielt, da dieser nach dem Einsetzen von Gl. (4.9) in Gl. (4.3) herausgekürzt werden kann. Daher soll hier ein allgemeiner Lösungsvektor $\tilde{\mathbf{F}}_{\mathrm{SNR}}(\Omega)$ eingeführt werden, welcher das SNR in Gl. (4.3) maximiert[2] und einen beliebigen komplexen Skalar $\zeta(\Omega)$ zulässt:

$$\tilde{\mathbf{F}}_{\mathrm{SNR}}(\Omega) = \zeta(\Omega)\mathbf{F}_{\mathrm{SNR}}(\Omega) \qquad \zeta(\Omega) \in \mathbb{C}. \tag{4.11}$$

Sicherstellung dieser Eigenschaft ein so genannter Regularisierungsterm eingefügt werden, siehe Kapitel 5.

[2] Aufgrund der Beziehung Gl. (4.11) lässt sich nicht von *dem* Eigenvektor, sondern vielmehr von *einem* Eigenvektor sprechen, der das Ausgangs-SNR maximiert.

Da im Falle der mehrkanaligen Sprachsignalverarbeitung in der Regel nicht die Sprachkomponente separat beobachtet werden kann, ist es auch nicht möglich die KLDS-Matrix $\mathbf{\Phi_{SS}}(\Omega)$ zu bestimmen. Daher kann zur Berechnung des gesuchten Eigenvektors nicht Gl. (4.4) herangezogen werden. Es kann jedoch die Störkomponente in Sprachpause-Sequenzen getrennt aufgenommen und somit die KLDS-Matrix $\mathbf{\Phi_{NN}}(\Omega)$ geschätzt werden. Zusätzlich kann zu Zeiten von Sprachaktivität Sprache-plus-Störung an den Mikrophonen beobachtet und folglich auch die KLDS-Matrix $\mathbf{\Phi_{XX}}(\Omega)$ geschätzt werden. Daher kann mit

$$\mathbf{\Phi_{XX}}(\Omega) = \mathbf{\Phi_{SS}}(\Omega) + \mathbf{\Phi_{NN}}(\Omega) \tag{4.12}$$

Gl. (4.1) umgeschrieben werden zu

$$\mathrm{SNR}_{\mathrm{Array}}(\Omega) = \frac{\mathbf{F}^H(\Omega)\mathbf{\Phi_{XX}}(\Omega)\mathbf{F}(\Omega)}{\mathbf{F}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}(\Omega)} - 1. \tag{4.13}$$

Für das Eigenwertproblem in Gl. (4.13) bezüglich der Matrizen $\mathbf{\Phi_{XX}}(\Omega)$ und $\mathbf{\Phi_{NN}}(\Omega)$ maximiert ebenfalls der Eigenvektoren $\mathbf{F}^{(\max)}(\Omega)$ bzw. $\mathbf{F}_{\mathrm{SNR}}(\Omega)$ den Rayleigh Quotienten, allerdings ergibt sich dann der zugehörige größte Eigenwert

$$\lambda_X^{(\max)}(\Omega) = \frac{{\mathbf{F}^{(\max)}}^H(\Omega)\mathbf{\Phi_{XX}}(\Omega)\mathbf{F}^{(\max)}(\Omega)}{{\mathbf{F}^{(\max)}}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}^{(\max)}(\Omega)} = \lambda_S^{(\max)}(\Omega) + 1. \tag{4.14}$$

Äquivalent zu Gl. (4.5) gilt hier nun die Eigenwertgleichung

$$\mathbf{\Phi_{XX}}(\Omega)\mathbf{F}(\Omega) = \lambda_X^{(\max)}(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}(\Omega), \tag{4.15}$$

welche nach vorheriger Bestimmung von $\mathbf{\Phi_{XX}}(\Omega)$ und $\mathbf{\Phi_{NN}}(\Omega)$ die Berechnung eines Koeffizientenvektors $\tilde{\mathbf{F}}_{\mathrm{SNR}}(\Omega)$ zulässt. Hierfür existieren eine Reihe von iterativen Lösungen [MRP96, GV99, Mor04, RPW04, YXYZ06, SK06], deren Eigenschaften eingehender im Kapitel 5 untersucht werden sollen; insbesondere im Zusammenhang mit dem akustischen *Beamforming* [WHU05, HUW05].

Array Gain Für den optimalen *Beamformer* nach Gl. (4.10) ergibt sich mit Gl. (3.47) ein SNR-Gewinn von

$$\begin{aligned} G_{\mathrm{SNR}}(\Omega) &= \frac{\mathbf{F}_{\mathrm{SNR}}^H(\Omega)\mathbf{\Phi_{SS}}(\Omega)\mathbf{F}_{\mathrm{SNR}}(\Omega)}{\mathbf{F}_{\mathrm{SNR}}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}_{\mathrm{SNR}}(\Omega)} \cdot \frac{\mathrm{Spur}\{\mathbf{\Phi_{NN}}(\Omega)\}}{\mathrm{Spur}\{\mathbf{\Phi_{SS}}(\Omega)\}} && (4.16)\\ &= \mathbf{H}^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega) \cdot \frac{\mathrm{Spur}\{\mathbf{\Phi_{NN}}(\Omega)\}}{\mathbf{H}^H(\Omega)\mathbf{H}(\Omega)}. && (4.17) \end{aligned}$$

White Noise Gain Der SNR-Gewinn bezüglich eines unkorrelierten Schallfeldes kann angegeben werden mit

$$G_{\mathrm{SNR}}^{\mathrm{W}}(\Omega) = M. \tag{4.18}$$

Beim Vergleich von Gl. (4.18) bzw. Gl. (4.17) und dem *White Noise Gain* des DSB in Gl. (3.55) ist zu erkennen, dass der Gewinn der Größenordnung M beim DSB nur erzielt wird, wenn die Ausbreitung des Sprachsignals im Freifeld angenommen wird. Hingegen ist der maximale Gewinn beim optimalen *Beamforming* Gl. (4.18) für beliebige Ausbreitungsbedingungen möglich.

Anmerkungen

Bei dem Vergleich zwischen dem einfachen *Beamforming*-Verfahren mittels DSB und einer mehrkanaligen Filterung mit den Koeffizienten $\tilde{\mathbf{F}}_{\mathrm{SNR}}(\Omega)$ fallen abgesehen von der Leistungsfähigkeit einige gravierende Unterschiede bezüglich der Berechnung der Filterkoeffizienten auf. Als Wissensquellen zur Bestimmung von $\tilde{\mathbf{F}}_{\mathrm{SNR}}(\Omega)$ über das Max-SNR-Kriterium sind lediglich die Schätzungen der KLDS-Matrizen[3] $\mathbf{\Phi_{XX}}(\Omega)$ und $\mathbf{\Phi_{NN}}(\Omega)$ notwendig. Es wird kein weiteres Wissen über die Sprecherrichtung θ_s und die Mikrophongeometrie (Positionen $\mathbf{p}_i$ bzw. Abstand d) benötigt. Soll jedoch als erste Verarbeitungseinheit eine Laufzeitkompensation erfolgen wie z. B. bei einem DSB, so sind dies zwingend notwendige Informationen. Zusätzlich ist bei einem realen System auf eine gleiche Verstärkung der eingehenden Mikrophonsignale zu achten, um eine kohärente Überlagerung zu gewährleisten. Bei einem DSB erfolgt dies über einen separaten Algorithmus zur Pegelanpassung, entweder im laufenden Betrieb oder während einer Kalibrierung in der Startphase [NCG01]. Da sich für die Maximierung des Ausgangs-SNR mittels $\tilde{\mathbf{F}}_{\mathrm{SNR}}(\Omega)$ automatisch pegelkompensierende Beträge für die Filterkoeffizienten ergeben, ist bei der Nutzung des Eigenvektors zum *Beamforming* eine separate Bestimmung der Eingangspegel nicht erforderlich.

Der entscheidende Nachteil bei der Nutzung des Eigenvektors $\tilde{\mathbf{F}}_{\mathrm{SNR}}(\Omega)$ zum akustischen *Beamforming* ergibt sich aufgrund der frequenzabhängigen Skalierung $\zeta(\Omega)$, die für jede betrachtete Spektralkomponente je nach Berechnungsvorschrift beliebig ausfallen kann. Dies bedeutet für die Verarbeitung von breitbandigen Sprachsignalen eine Verzerrung des Nutzsignals, obschon für jede Spektralkomponente das Ausgangs-SNR maximal ist. Auf diese Problematik sowie Lösungsvorschläge zur automatischen Kontrolle des Effekts wird in Kapitel 6 detailliert eingegangen.

4.2 Minimum Variance

Der nächste Ansatz zur Herleitung optimaler Filterkoeffizienten beruht auf der Minimierung der Störvarianz. Dazu wird

$$\mathbf{F}^H(\Omega)\mathbf{\Phi_{XX}}(\Omega)\mathbf{F}(\Omega) = \phi_{S_cS_c}(\Omega)\mathbf{F}^H(\Omega)\mathbf{H}(\Omega)\mathbf{H}^H(\Omega)\mathbf{F}(\Omega) + \mathbf{F}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}(\Omega) \tag{4.19}$$

betrachtet. Da mit $\mathbf{H}(\Omega)$ die Raumübertragungsfunktion zwischen dem Sprecher und der Mikrophongruppe bezeichnet ist, kann das Skalarprodukt $\mathbf{F}^H(\Omega)\mathbf{H}(\Omega)$ als gesamte Übertragungsfunktion zwischen Sprecher und dem Ausgang des *Arrays* interpretiert werden. Nun soll für genau diese gemeinsame Übertragungsfunktion folgende lineare Bedingung (engl. *Linear Constraint*) gelten

$$\mathbf{F}^H(\Omega)\mathbf{H}(\Omega) = W(\Omega). \tag{4.20}$$

Ausgehend von Gl. (4.19) kann mit der spektralen Gewichtung[4] $W(\Omega)$ des Quellensignals aus Gl. (4.20) die Kostenfunktion

$$J_{\mathrm{MV}}(\mathbf{F}(\Omega)) = \mathbf{F}^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}(\Omega) + \Re\{\beta^*(\Omega)(W(\Omega) - \mathbf{F}^H(\Omega)\mathbf{H}(\Omega))\} \tag{4.21}$$

[3] Es soll hier erwähnt werden, dass zur Schätzung der KLDS-Matrizen eine zusätzliche Informationsquelle in Form einer Sprache/Pause-Detektion vorausgesetzt wird. Allerdings ist solch eine Unterteilung der Eingangsdaten in Sprache- und Pausesequenzen ebenfalls zur Schätzung der Sprecherrichtung nötig.

[4] Für das gefilterte Sprachsignal ergibt sich am *Beamformer*-Ausgang $\mathbf{F}^H(\Omega)\mathbf{S}(\Omega) = \mathbf{F}^H(\Omega)S_c(\Omega)\mathbf{H}(\Omega) = S_c(\Omega)W(\Omega)$, also das mit $W(\Omega)$ gewichtete Quellensignal. Mittels dieser Nebenbedingung können z. B. Spektralkomponenten in denen *a priori* keine oder wenige Sprachanteile vorhanden sind gedämpft werden (Bandpass).

aufgestellt und minimiert werden. In Gl. (4.21) ist mit $\Re\{\cdot\}$ die Realteilbildung und mit $\beta(\Omega)$ der frequenzabhängige Lagrange-Multiplikator[5] bezeichnet. Der Methode nach Lagrange folgend [Hay02] wird der Gradient

$$\nabla_{\mathbf{F}} J_{\mathrm{MV}}(\mathbf{F}(\Omega)) = 2\frac{\partial J_{\mathrm{MV}}(\mathbf{F}(\Omega))}{\partial \mathbf{F}^*} = \mathbf{\Phi_{NN}}(\Omega)\mathbf{F}(\Omega) - \beta^*(\Omega)\mathbf{H}(\Omega) \tag{4.22}$$

zu Null gesetzt, so dass sich mit

$$\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}(\Omega) = \beta^*(\Omega)\mathbf{H}(\Omega) \tag{4.23}$$

der unbekannte Lagrange-Multiplikator zu

$$\beta(\Omega) = \frac{\mathbf{F}^H(\Omega)\mathbf{H}(\Omega)}{\mathbf{H}^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega)}. \tag{4.24}$$

ergibt. Weiter wird für die optimalen Filterkoeffizienten angenommen, dass die Bedingung Gl. (4.20) eingehalten wird,

$$\beta(\Omega) = \frac{W(\Omega)}{\mathbf{H}^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega)}, \tag{4.25}$$

und letztendlich der optimale Koeffizientenvektor nach Einsetzen von Gl. (4.25) in Gl. (4.23) berechnet werden kann:

$$\mathbf{F}_{\mathrm{GMV}}(\Omega) = W^*(\Omega)\frac{\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega)}{\mathbf{H}^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega)}. \tag{4.26}$$

Mit dem Index "GMV" soll auf die verallgemeinerte Minimierung der Varianz (engl. *Generalized Minimum Variance*, GMV) bezüglich der Störung hingewiesen werden, wobei die Verallgemeinerung auf die Verwendung der kompletten Raumübertragungsfunktion $\mathbf{H}(\Omega)$ zurückzuführen ist.

Für die Forderung eines unverzerrt gebliebenen Sprachsignals am *Beamformer*-Ausgang ist die Bedingung Gl. (4.20) für alle Frequenzen konstant zu setzen

$$W(\Omega) = 1. \tag{4.27}$$

Dadurch ergibt sich ein *Beamformer* mit einer unverzerrten Antwort (engl. *Distortionless Response*, DR) bezüglich des Sprachsignals und der damit verbundene Koeffizientenvektor

$$\mathbf{F}_{\mathrm{GMVDR}}(\Omega) = \frac{\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega)}{\mathbf{H}^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega)}. \tag{4.28}$$

Es kann leicht überprüft werden, dass das mit $\mathbf{F}_{\mathrm{GMVDR}}(\Omega)$ gefilterte Sprachsignal am Ausgang des *Beamformers* dem unverzerrten Quellensignal entspricht:

$$\mathbf{F}_{\mathrm{GMVDR}}^H(\Omega)\mathbf{S}(\Omega) = \frac{\mathbf{H}^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)}{\mathbf{H}^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega)}\mathbf{H}(\Omega)S_c(\Omega) = S_c(\Omega). \tag{4.29}$$

[5]Entgegen der üblichen Notation λ für den Lagrange-Multiplikator soll hier die Bezeichnung β verwendet werden um Verwechslungen mit der Kennzeichnung von Eigenwerten zu vermeiden.

Vergleicht man nun Gl. (4.28) mit der Max-SNR-Lösung Gl. (4.10), so kann folgender Zusammenhang festgestellt werden:

$$\begin{aligned} \mathbf{F}_{\mathrm{GMVDR}}(\Omega) &= w_{\mathrm{GMVDR}} \boldsymbol{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega) && (4.30)\\ &= w_{\mathrm{GMVDR}}(\Omega)\mathbf{F}_{\mathrm{SNR}}(\Omega) && (4.31) \end{aligned}$$

mit dem skalaren Faktor

$$w_{\mathrm{GMVDR}}(\Omega) = \frac{1}{\mathbf{H}^H(\Omega)\boldsymbol{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega)}. \tag{4.32}$$

Im Gegensatz zur Berechnung von $\tilde{\mathbf{F}}_{\mathrm{SNR}}(\Omega)$ in Gl. (4.11) über die Eigenwertzerlegung bezüglich $\boldsymbol{\Phi}_{\mathbf{NN}}(\Omega)$ und $\boldsymbol{\Phi}_{\mathbf{XX}}(\Omega)$ muss zur Bestimmung von $\mathbf{F}_{\mathrm{GMVDR}}(\Omega)$ in Gl. (4.28) die Raumübertragungsfunktion $\mathbf{H}(\Omega)$ bekannt sein (aufgrund der Gewichtung Gl. (4.32)). Da dies in der Regel nicht der Fall ist, werden üblicherweise nur die Verzögerungen der direkten Ausbreitungspfade zwischen den Mikrophonen und dem Sprecher geschätzt und $\mathbf{H}(\Omega)$ wird in Gl. (4.28) durch den *Steering Vector* Gl. (3.30) ersetzt

$$\mathbf{F}_{\mathrm{MVDR}}(\Omega) = \frac{\boldsymbol{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{d}(\Omega,\theta_t)}{\mathbf{d}^H(\Omega,\theta_t)\boldsymbol{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{d}(\Omega,\theta_t)}. \tag{4.33}$$

Das resultierende $\mathbf{F}_{\mathrm{MVDR}}(\Omega)$ in Gl. (4.33) ist unter dem Begriff *Minimum Variance Distortionless Response Beamformer* bekannt und wird daher mit dem Index "MVDR" bezeichnet. Zu beachten ist hierbei, dass beim Übergang von Gl. (4.28) nach Gl. (4.33) der Parameter θ_t aufgeführt wird. Aufgrund der Definition Gl. (3.16) gilt die verkürzte Schreibweise $\mathbf{H}(\Omega)$ für die Raumübertragungsfunktion, obschon sie von der Position der Schallquelle $\mathbf{p}_s$ und der Mikrophone $\mathbf{p}_i$, $i = 1, \ldots, M$ abhängt. Da bei einem linear angeordneten *Array* für $\mathbf{F}_{\mathrm{MVDR}}(\Omega)$ eine Sprecherrichtungsschätzung θ_t notwendig ist, und diese nicht zwangsläufig identisch mit der wahren Richtung θ_s übereinstimmen muss, wird der Parameter θ_t im *Steering Vector* beibehalten.

Für den GMVDR *Beamformer* ergibt sich offensichtlich der gleiche SNR-Gewinn Gl. (4.17) wie für den optimalen *Beamformer* $\mathbf{F}_{\mathrm{SNR}}(\Omega)$. Hingegen stellen sich je nach Raumsituation und Störgeräuschfeld beim MVDR *Beamformer* geringe Unterschiede im Vergleich zur verallgemeinerten Lösung ein. Diese Unterschiede sollen im Abschnitt 4.6 analysiert werden.

Anmerkungen

In der Literatur zum akustischen *Beamforming* wird nur vereinzelt auf die konstruktive Nutzung der Mehrwegeausbreitung eingegangen [NNS01, KHJ06] und fast ausschließlich die Minimierung der Ausgangsleistung des *Beamformers* mit der Nebenbedingung eines unverzerrten Signals aus der *Look Direction* als Optimierungskriterum herangezogen. Dabei stellt insbesondere die adaptive Lösung nach Frost [Fro72] eine immer noch stark verbreitete Basis dar. Da der MVDR *Beamformer* stark von der genauen Schätzung der Richtung des gewünschten Quellensignals, also von der Bestimmung des *Steering Vectors* abhängt, beschäftigt sich eine Vielzahl von Veröffentlichungen zu adaptiven MVDR *Beamformern* mit Robustheitsaspekten [LS05, HGJ06, JHLCCC06].

Eine Realisierung des MVDR *Beamformers* mit der Optimierung hinsichtlich der Direktivität nimmt für zahlreiche Autoren einen besonderen Stellenwert ein [Täg98, BSK99a, Elk00, JG00, BS01]. Diese superdirektiven *Beamformer* werden für den Fall eines diffusen

Störschallfelds optimiert; es wird also für $\mathbf{\Phi_{NN}}(\Omega)$ *a priori* die Kohärenz-Matrix des diffusen Störschallfelds eingesetzt. Dabei ist jedoch auf die Besonderheit der Verstärkung von räumlich unkorreliertem Rauschen zu achten [Bit02]. Die Adaption ist dann auf die Bestimmung der Sprecherrichtung konzentriert.

4.3 Maximum Likelihood

Für den *Maximum-Likelihood*-Ansatz wird davon ausgegangen, dass das Quellensignal $S_c(\Omega)$ und das Rauschen am i-ten Mikrophon $N_i(\Omega)$ mittelwertfreie, komplexe, gaußverteilte Zufallsvariablen sind. Weiterhin sollen $S_c(\Omega_k)$ und $N_i(\Omega_k)$ der Frequenz Ω_k jeweils statistisch unabhängig von $S_c(\Omega_\nu)$ und $N_i(\Omega_\nu)$ für unterschiedliche Frequenzen $\Omega_k \neq \Omega_\nu$ sein. Mit Hilfe dieser Voraussetzungen kann die *a posteriori* Wahrscheinlichkeitsdichtefunktion (engl. *Probability Density Function*, PDF)

$$p(Y(\Omega)|S_c(\Omega)) = \eta(\Omega)e^{-\hat{\mathbf{N}}^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\hat{\mathbf{N}}(\Omega)} \tag{4.34}$$

angegeben und als *Likelihood* aufgefasst werden [Lev64]; mit der Schätzung für das Rauschen

$$\hat{\mathbf{N}}(\Omega) = \mathbf{X}(\Omega) - S_c(\Omega)\mathbf{H}(\Omega) \tag{4.35}$$

und der skalaren Konstante $\eta(\Omega)$, welche unabhängig von $S_c(\Omega)$ ist. Somit ergibt sich die zu minimierende negative *Log-Likelihood*-Funktion

$$\mathcal{L}(\mathbf{X}(\Omega)) = \tilde{\eta}(\Omega)\hat{\mathbf{N}}^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\hat{\mathbf{N}}(\Omega). \tag{4.36}$$

Durch null setzen der partiellen Ableitung von $\mathcal{L}(\mathbf{X}(\Omega))$ nach $S_c(\Omega)$ erhält man schließlich die Schätzung $\hat{S}_c(\Omega)$ für das Quellensignal, welches die *Log-Likelihood*-Funktion Gl. (4.36) minimiert

$$\hat{S}_c(\Omega) = \frac{\mathbf{H}^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)}{\mathbf{H}^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega)}\mathbf{X}(\Omega) = \mathbf{F}_{\mathrm{GML}}^H(\Omega)\mathbf{X}(\Omega). \tag{4.37}$$

In Gl. (4.37) ist mit $\mathbf{F}_{\mathrm{GML}}(\Omega)$ der Koeffizientenvektor des verallgemeinerten ML-Ansatzes (engl. *Generalized Maximum Likelihood*, GML) bezeichnet

$$\mathbf{F}_{\mathrm{GML}}(\Omega) = \frac{\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega)}{\mathbf{H}^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega)} = \mathbf{F}_{\mathrm{GMVDR}}(\Omega), \tag{4.38}$$

welcher identisch mit der GMVDR-Lösung Gl. (4.28) ist. Auch hier soll vollständigkeitshalber noch die vereinfachte Variante

$$\mathbf{F}_{\mathrm{ML}}(\Omega) = \frac{\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{d}(\Omega,\theta_t)}{\mathbf{d}^H(\Omega,\theta_t)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{d}(\Omega,\theta_t)} = \mathbf{F}_{\mathrm{MVDR}}(\Omega) \tag{4.39}$$

mit dem *Steering Vektor* $\mathbf{d}(\Omega,\theta_t)$ angegeben sein.

Anmerkungen

Die MVDR-Filterkoeffizienten stellen also den *Maximum-Likelihood*-Schätzer für das Quellensignal dar, wenn die Sprecherrichtung und die KLDS-Matrix der Störung bekannt sind. Alternativ wird in [VSO97] von keinerlei Wissen über die Störung ausgegangen, sondern von

Annahmen bezüglich des Nutzsignals. Da aber der Zusammenhang Gl. (4.38) bzw. Gl. (4.39) besteht, werden deutlich weniger *Maximum-Likelihood-Beamforming*-Verfahren im Vergleich zur MVDR-Lösung in der Literatur diskutiert. Ein großer Teil beschäftigt sich mit Robustheitsaspekten und dem Einfluss einer fehlerhaften Richtungsschätzung [LS05].

Interessant ist das in [DCP03] entwickelte ML-Verfahren, welches insbesondere auf das Problem verhallter Signale eingeht. Dort wird ein so genannter *Maximum Likelihood Steered Adaptive Beamformer* beschrieben, in dem ein stark nichtlinearer ML-Ansatz mit Hilfe eines modifizierten Newton Adaptionsalgorithmus ohne Nebenbedingung gelöst wird und zur deutlichen Reduzierung von Störinterferenzen führt.

Bei dem in [SRS04, BSRG05] vorgestellten Verfahren steht die Anwendung eines *Maximum Likelihood Beamformers* zur Reduzierung der Wortfehlerrate eines nachgeschalteten Spracherkenners im Vordergrund. Dabei werden die *Beamformer*-Koeffizienten bezüglich eines ML-Kriteriums optimiert, in welches die Parameter des vorläufigen Erkennungsergebnisses des Spracherkenners einfließen. Der *Beamformer* wird dann derat adaptiert, dass die Wahrscheinlichkeit dafür steigt, dass die iterativ erkannte Wortfolge mit der gesprochenen Sequenz übereinstimmt.

4.4 Minimum Mean Squared Error

Zunächst soll davon ausgegangen werden, dass das gewünschte Quellensignal $S_c(\Omega)$ bekannt sei und sich somit folgender Ausdruck für den mittleren quadratischen Fehler (engl. *Mean Squared Error*, MSE) angeben lässt:

$$\begin{aligned} J_{\text{MSE}}(\mathbf{F}(\Omega)) &= E\{|S_c(\Omega) - \mathbf{F}^H(\Omega)\mathbf{X}(\Omega)|^2\} && (4.40)\\ &= \phi_{S_cS_c} - \mathbf{F}^H(\Omega)\boldsymbol{\phi}_{\mathbf{X}S_c}(\Omega) - \boldsymbol{\phi}^H_{\mathbf{X}S_c}(\Omega)\mathbf{F}(\Omega) + \mathbf{F}^H(\Omega)\boldsymbol{\Phi}_{\mathbf{XX}}(\Omega)\mathbf{F}(\Omega). && (4.41)\end{aligned}$$

Zur Minimierung des mittleren quadratischen Fehlers (engl. *Minimum* MSE, MMSE) wird der Gradient

$$\nabla_{\mathbf{F}} J_{\text{MSE}}(\mathbf{F}(\Omega)) = -2\boldsymbol{\phi}_{\mathbf{X}S_c}(\Omega) + 2\boldsymbol{\Phi}_{\mathbf{XX}}(\Omega)\mathbf{F}(\Omega) \tag{4.42}$$

zu Null gesetzt und liefert die verallgemeinerte MMSE-Lösung (GMMSE)

$$\mathbf{F}_{\text{GMMSE}}(\Omega) = \boldsymbol{\Phi}^{-1}_{\mathbf{XX}}(\Omega)\boldsymbol{\phi}_{\mathbf{X}S_c}(\Omega), \tag{4.43}$$

unter der Voraussetzung, dass $\boldsymbol{\Phi}_{\mathbf{XX}}(\Omega)$ nicht singulär und somit invertierbar ist. Gl. (4.43) ist die Wiener-Hopf-Gleichung in Matrix-Form und kann daher als mehrkanaliges Wiener Filter (engl. *Multi Channel Wiener Filter*, MWF) gesehen werden. Mit der additiven Zusammensetzung

$$\boldsymbol{\Phi}_{\mathbf{XX}}(\Omega) = \phi_{S_cS_c}(\Omega)\mathbf{H}(\Omega)\mathbf{H}^H(\Omega) + \boldsymbol{\Phi}_{\mathbf{NN}}(\Omega) \tag{4.44}$$

und dem Matrix Inversion Lemma (siehe Anhang A.2), ist es möglich das Wiener Filter Gl. (4.43) in die faktorisierte Form

$$\mathbf{F}_{\text{GMMSE}}(\Omega) = w_{\text{WPF}}(\Omega)\mathbf{F}_{\text{GMVDR}}(\Omega). \tag{4.45}$$

zu überführen. Der skalare Faktor

$$w_{\text{WPF}}(\Omega) = \frac{\phi_{S_cS_c}(\Omega)}{\phi_{S_cS_c}(\Omega) + \left(\mathbf{H}^H(\Omega)\boldsymbol{\Phi}^{-1}_{\mathbf{NN}}(\Omega)\mathbf{H}(\Omega)\right)^{-1}} \tag{4.46}$$

kann als frequenzabhängige Nachfilterung (engl. *Wiener Post Filter*, WPF) interpretiert werden [SBM01]. Dies wird um so deutlicher, wenn das Leistungsdichtespektrum der Störung $\phi_{N_oN_o}(\Omega)$ am Ausgang des GMVDR *Beamformers* betrachtet wird:

$$\begin{aligned}\phi_{N_oN_o}(\Omega) &= \mathbf{F}^H_{\mathrm{GMVDR}}(\Omega)\mathbf{\Phi_{NN}}(\Omega)\mathbf{F}_{\mathrm{GMVDR}}(\Omega) &(4.47)\\ &= \left(\mathbf{H}^H(\Omega)\mathbf{\Phi}^{-1}_{\mathbf{NN}}(\Omega)\mathbf{H}(\Omega)\right)^{-1} &(4.48)\\ &= w_{\mathrm{GMVDR}}(\Omega). &(4.49)\end{aligned}$$

Somit ergibt sich für die Nachfilterung Gl. (4.46) der Ausdruck

$$w_{\mathrm{WPF}}(\Omega) = \frac{\phi_{S_cS_c}(\Omega)}{\phi_{S_cS_c}(\Omega) + \phi_{N_oN_o}(\Omega)}. \tag{4.50}$$

Wie bereits gezeigt, maximiert die GMVDR-Lösung in gleicher Weise wie die Max-SNR-Lösung zwar das Schmalband-SNR[6], aber nicht zwangsläufig das Breitband-SNR. Dies wird erst durch das nachgeschaltete mehrkanalige Wiener Filter $w_{\mathrm{WPF}}(\Omega)$ erreicht. Diese Eigenschaft ist sehr gut an dem nachgeschalteten Wiener Filter Gl. (4.50) zu erkennen. Während mit den GMVDR-Filterkoeffizienten die räumliche Information ausgenutzt wird und das Signal in Blickrichtung unverzerrt erhalten bleibt, erfolgt eine spektrale Dämpfung durch $w_{\mathrm{WPF}}(\Omega)$ für Frequenzkomponenten mit einem geringen SNR. Dadurch wird zwar eine Verzerrung des Sprachsignals[7] in Kauf genommen, aber eben auch eine SNR-Maximierung des breitbandigen Audiosignals erzielt.

Für die optimalen Filterkoeffizienten $\mathbf{F}_{\mathrm{GMMSE}}(\Omega)$ kann nun wieder äquivalent zu Gl. (4.31) ein direkter, skalarer Zusammenhang zwischen der MMSE- und der Max-SNR-Lösung angegeben werden:

$$\begin{aligned}\mathbf{F}_{\mathrm{GMMSE}}(\Omega) &= w_{WPF}(\Omega)w_{\mathrm{GMVDR}}(\Omega)\mathbf{F}_{\mathrm{SNR}}(\Omega) &(4.51)\\ &= w_{\mathrm{GMMSE}}(\Omega)\mathbf{F}_{\mathrm{SNR}}(\Omega), &(4.52)\end{aligned}$$

mit Gl. (4.49) kann das Nachfilter in kompakter Schreibweise zu

$$w_{\mathrm{GMMSE}}(\Omega) = \frac{\phi_{S_cS_c}(\Omega)w_{\mathrm{GMVDR}}(\Omega)}{\phi_{S_cS_c}(\Omega) + w_{\mathrm{GMVDR}}(\Omega)}. \tag{4.53}$$

formuliert werden.

Wie auch beim MV-Ansatz wird bei der Realisierung von MMSE *Beamformern* nach dem oben beschriebenen Schema die Raumübertragungsfunktion durch den *Steering Vector* ersetzt. Dadurch ergibt sich die in der Literatur [SBM01] übliche Variante

$$\mathbf{F}_{\mathrm{MMSE}}(\Omega) = \frac{\phi_{S_cS_c}(\Omega)}{\phi_{S_cS_c}(\Omega) + \left(\mathbf{d}^H(\Omega)\mathbf{\Phi}^{-1}_{\mathbf{NN}}(\Omega)\mathbf{d}(\Omega)\right)^{-1}}\mathbf{F}_{\mathrm{MVDR}}(\Omega) \tag{4.54}$$

für die Filterkoeffizienten. Für den Fall, dass das SNR am Eingang des *Arrays* hoch ist, liefern also offensichtlich MVDR und MMSE[8] *Beamformer* sehr ähnliche Ergebnisse, wie bereits in [Gri67] untersucht wurde.

[6] Das Breitband-SNR bezeichnet das SNR bestimmt über alle enthaltenen Frequenzkomponenten. Hingegen ist das Schmalband-SNR das frequenzabhängige SNR.

[7] In [RBB03, ZHA04] werden Methoden zur Minimierung der Sprachsignalverzerrung durch psychoakustisch motivierte Maskierungseffekte beschrieben.

[8] Für die Annahme gaußverteilter Real- und Imaginäranteile der frequenzabhängigen Sprach- und Störsignale ist der optimale MMSE-Schätzer identisch mit dem *maximum a posteriori* (MAP) Schätzer [VT68].

Anmerkungen

Die offensichtliche Schwierigkeit zur Berechnung der MMSE-Filterkoeffizienten besteht in der Schätzung des Sprachsignals, oder, allgemeiner ausgedrückt, in dem Problem der Erzeugung eines Referenzsignals. Bei der Anwendung von *Beamforming*-Verfahren für Antennen-*Arrays* ist es möglich, ein Pilot-Signal aus der *Look Direction* als Referenzsignal zu nutzen. In [WMGG67] ist bereits solch eine Methode inklusive Adaption mit Hilfe der kleinsten Fehlerquadrate vorgestellt. Äquivalent dazu werden in [NCG01] über eine Kalibrierungs-Sprachsequenz die optimalen Filterkoeffizienten für die Mikrophongruppe in einem Kraftfahrzeug berechnet. Dabei beinhaltet die MMSE-Schätzung repräsentative Einflüsse der verwendeten *Hardware* sowie der Mikrophon- und Sprecherposition, siehe auch [GN02, NGL05].

Die populärste MMSE-Variante zur Umsetzung von Gl. (4.54) beruht auf der Annahme von unkorrelierten Störsignalen in den einzelnen Mikrophonpfaden. Dann kann eine Mittelung der Kreuzleistungsdichten zwischen jeweils zwei Signalpaaren zur Schätzung $\phi_{S_cS_c}(\Omega)$ hergenommen werden [Zel88]. Da jedoch diese Annahme für ein gerichtetes oder diffuses Störschallfeld nicht bzw. nur bedingt für einen bestimmten Frequenzbereich gilt (vgl. Abschnitt 2.4), ist eine Verbesserung der Schätzung durch *a priori* Annahmen für die räumliche Korrelation des Störgeräuschfeldes in [SW92, MMU98, BSK99b] und durch explizite Berechnung in [MB02, MB03] mit berücksichtigt worden.

Eine andere Variante ergibt sich durch die statistische Auswertung der durch das Sprach- und Störsignal aufgespannten Unterräume der Kovarianzmatrizen[9] im Zeitbereich. Dabei ergeben sich Filterkoeffizienten aus Eigenvektoren mittels einer verallgemeinerten Singulärwertzerlegung [DM01, SMW02]. Entstehende Sprachverzerrungen werden in [DSWM05, CBHD06] geschätzt und konstruktiv für die Adaption benutzt.

4.5 Experimente zur verallgemeinerten Lösung

Im folgenden Abschnitt werden einige Ergebnisse zur experimentellen Untersuchung des verallgemeinerten GMVDR-Ansatzes Gl. (4.28) präsentiert. Hierfür wurden die Anordnungen *Szenario-1* und *Szenario-2* aus dem Anhang C verwendet (also eine Sprecherrichtung von $\theta_s = \theta_t = 45°$) und die Übertragungsfunktion zwischen der Sprecherposition und den Sensoren mit Hilfe der reinen Sprachdaten geschätzt. Die Schätzung der Übertragungsfunktion erfolgte durch den Algorithmus 3 (S-Grad-GG) aus Abschnitt 5.1.5 mit der Normalisierung aus Abschnitt 6.1. Für den Fall von $M = 5$ Sensoren, einer Nachhallzeit von $T_{60} = 0{,}1\,\mathrm{s}$ und einer DFT-Länge von $L = 256$ ergeben sich die in Bild 4.1 dargestellten Verläufe[10] für die erste und fünfte Raumimpulsantwort $h_1(n)$ und $h_5(n)$, sowie deren Schätzung $\hat{h}_1(n)$ und $\hat{h}_5(n)$. An den identifizierten Impulsantworten in Bild 4.1 sind nun zwei markante Eigenschaften zu erkennen. Zum einen ist ein Versatz des Anteiles, welcher zum direkten Pfad korrespondiert, um 4 Abtastwerte ($\sin(\theta_s) \cdot d \cdot f_{Ab} \cdot c^{-1} \cdot (M-1) = 4$) festzustellen. Und zum anderen können Anteile aufgrund von Reflexionen direkt den vorgegebenen Raumimpulsantworten zugeordnet werden.

[9] Aufgrund der Nichtstationarität der Sprache gelten die Annahmen bezüglich Stationarität und Unabhängigkeit der einzelnen Komponenten untereinander im Frequenzbereich nur näherungsweise. In [Her04] wird daher eine konsequente Herleitung optimaler Filterkoeffizienten über die Methode der kleinsten Fehlerquadrate (eng. *Least Squares Error*, LSE) im Zeitbereich durchgeführt.

[10] Zur besseren Darstellung in Bild 4.1 wurde lediglich der minimalphasige Anteil der Raumimpulsantworten verwendet [NA79].

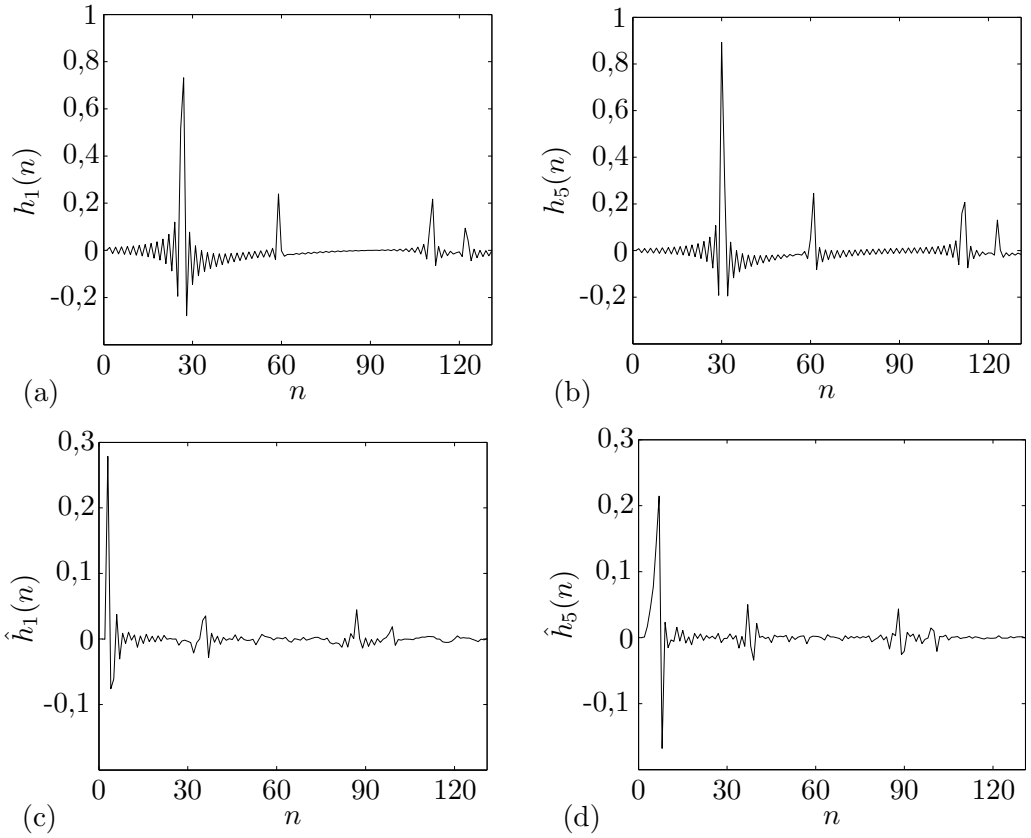

Bild 4.1: (a) Ausschnitt der ersten und (b) der fünften simulierten Raumimpulsantwort.
(c) Geschätzte erste und (d) geschätzte fünfte Raumimpulsantwort.

Als nächstes soll die Energiabfallkurve Gl. (2.13) untersucht werden, die für eine zeitdiskrete Impulsantwort geschrieben werden kann als

$$E_A(j) := -10 \log_{10} \frac{\sum_{n=j}^{\infty} h^2(n)}{\sum_{n=0}^{\infty} h^2(n)} \text{ dB}, \tag{4.55}$$

wobei $h(n)$ nun für drei Fälle betrachtet werden soll:

- Raumimpulsantwort "RIA": Die Raumimpulsantwort zwischen dem Sprecher und dem ersten Mikrophon.
- *Delay-and-Sum* "DS": Die kohärente Überlagerung (bezüglich des direkten Pfades) aller M Raumimpulsantworten.
- *Filter-and-Sum* "FS": Die gesamte Impulsantwort zwischen dem Sprecher und der Faltung mit den geschätzten Raumimpulsantworten: $h(j) = \sum_{i=1}^{M} h_i(n) \star \hat{h}_i(L-n)$.

Die Ergebnisse der Energiabfallkurven sind in Bild 4.2 über der Zeit aufgetragen. Es ist zu erkennen, dass zwar der Abfall der Kurven näherungsweise gleich ist, aber die konstruktive Überlagerung der direkten Ausbreitungspfade der Raumimpulsantworten führt zu einem größeren Sprung beim Maximum n_0 der gesamten Impulsantworten der DS- und FS-Kurven.

Dieses Verhalten resultiert in einem höheren Deutlichkeitsmaß (vgl. Gl. (2.11))

$$C_{50} = 10\log_{10}\frac{\sum_{n=n_0}^{n_0+n_{50}} h^2(n)}{\sum_{n=n_0+n_{50}+1}^{\infty} h^2(n)}, \tag{4.56}$$

wobei hier gilt $n_{50} = 50\,\mathrm{ms} \cdot f_{Ab} = 600$. Aufgrund der Verdeckung (vgl. 2.2) ist vor allem der Anfangsteil der Energiabfallkurve von besonderer Bedeutung. Die Anfangsnachhallzeit T_A für eine zeitdiskrete Impulsantwort ergibt sich äquivalent zu Gl. (2.12) als

$$-10\,\mathrm{dB} \stackrel{!}{=} 10\log_{10}\frac{\sum_{n=n_0}^{n_0+n_A} h^2(n)}{\sum_{n=n_0}^{\infty} h^2(n)}\ \mathrm{dB}, \tag{4.57}$$

mit $T_A = n_A/f_{Ab}$, und dem ersten Abtastwert n_A, für den Gl. (4.57) zutrifft. Die Ergebnisse für das Deutlichkeitsmaß und die Anfangsnachhallzeit sind in Bild 4.3 für variierende Werte der Länge B der geschätzten Raumimpulsantworten, der Mikrophonanzahl M und der Nachhallzeit T_{60} dargestellt.

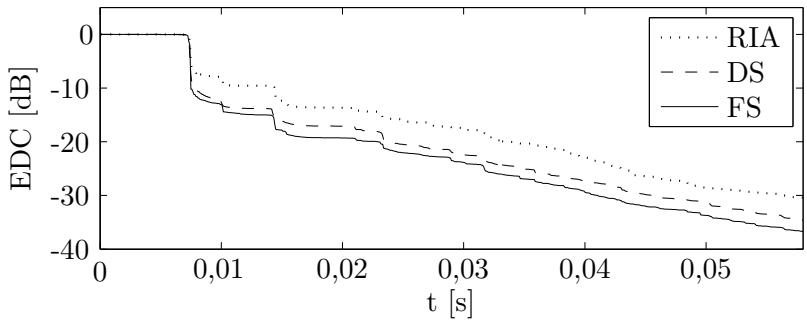

Bild 4.2: Energiabfallkurven für die erste Raumimpulsantwort (RIA), sowie die resultierenden Impulsantworten mittels kohärenter Überlagerung (DS) und der Filterung (FS) mit den geschätzten Raumimpulsantworten für $M = 5$ und $T_{60} = 0,1$ s.

Für die identifizierte Länge der Raumimpulsantworten in Bild 4.3 ist zu beachten, dass jeweils die Länge der DFT auf $L = 2B$ gesetzt wurde. Weiterhin soll angemerkt sein, dass in den Bildern der linken Spalte von 4.3 das Deutlichkeitsmaß und die Anfangsnachhallzeit für RIA und DS zum Vergleich eingetragen sind, obschon sie nicht von dem Parameter B abhängen.

Grundsätzlich kann an den Verläufen in Bild 4.3 festgestellt werden, dass durch die Faltung und Aufsummierung (FS) höhere Werte für das Deutlichkeitsmaß erzielt werden und ein schnellerer Abfall der Energiabfallkurve um 10 dB – gekennzeichnet durch die Anfangsnachhallzeit – im Vergleich zu RIA und DS erfolgt. Weiterhin ist offensichtlich, dass eine steigende Anzahl von Mikrophonen zu einem steigenden C_{50} und abfallendem T_A bei gleicher Nachhallzeit für DS und FS führt. Bei steigendem Nachhall sind die Verläufe aller Kurven ebenfalls folgerichtig, da der Sprung nach dem Anteil der EDC, der auf den direkten Pfad zurückzuführen ist, mit steigendem T_{60} deutlich kleiner und der anschließende lineare Abfall wesentlich geringer wird.

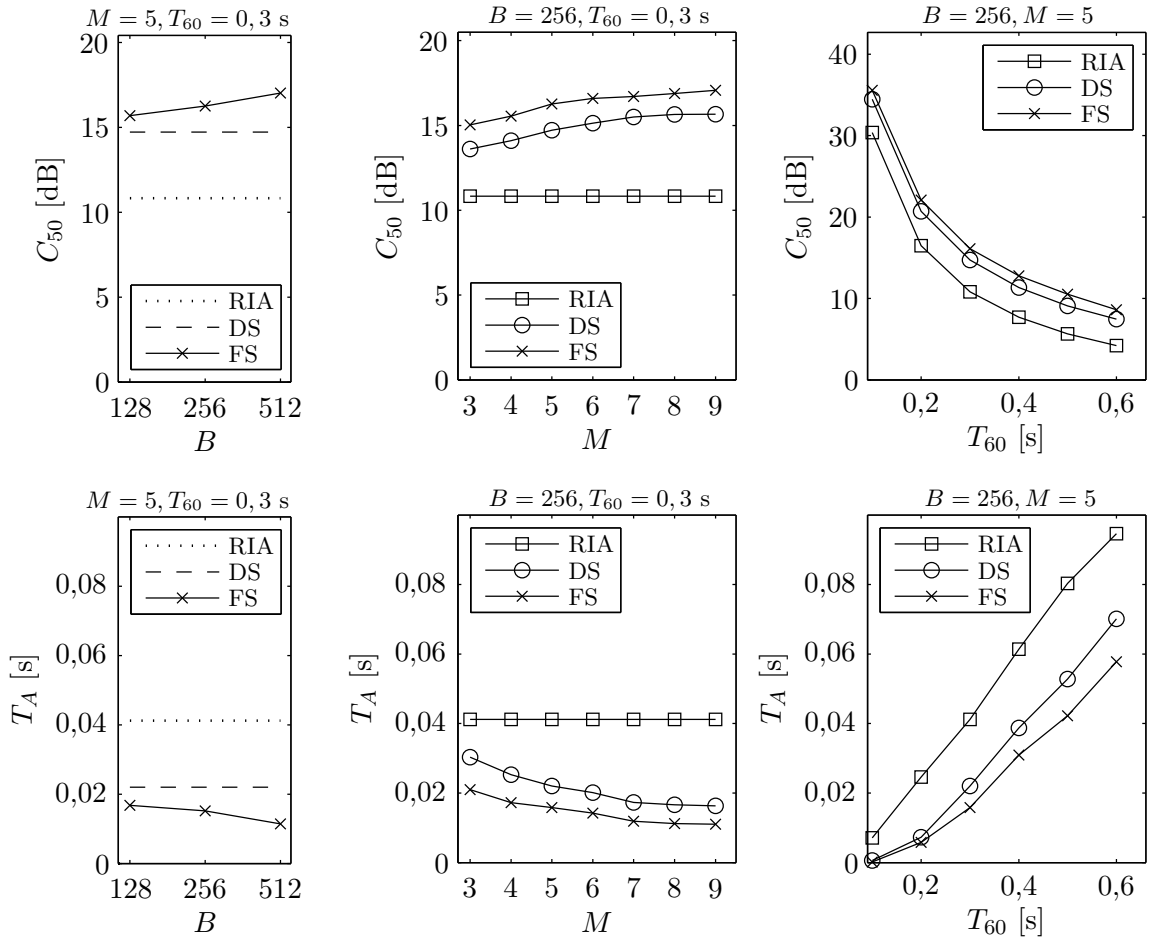

Bild 4.3: Auswertung des Deutlichkeitsmaßes C_{50} in der oberen Reihe und der Anfangsnachhallzeit T_A in der unteren Reihe für variierende Werte folgender Parameter: Länge B der geschätzten Raumimpulsantworten, Mikrophonanzahl M und Nachhallzeit T_{60}.

Nach den exemplarischen Auswertungen der Energiabfallkurven stellt sich die Frage, wie sich die Identifikation der Raumübertragungsfunktionen auf die Leistungsfähigkeit des GMVDR *Beamformers* auswirkt, also auf den Vergleich von Gl. (4.28) zu Gl. (4.33). Grundlage sind hier wieder *Szenario-1* und *Szenario-2*, wobei die KLDS-Matrix[11] der Störung durch eine Schätzung über $L = 512$ Werte, einem Vorschub von $B = L/2$ und einer Hann-Fensterung erfolgte. Die Inverse ist optimal bestimmt worden. Das Bild 4.4 (a) zeigt den SNR-Gewinn für den Fall, wenn die Störung nur aus weißem, unkorrelierten Rauschen besteht und in Bild 4.4 (b) ist der SNR-Gewinn für das gerichtete Tiefpassrauschen aus der Richtung $\theta_n = -20°$ dargestellt; jeweils für $M = 5$ Mikrophone aufgetragen über der Nachhallzeit. Für das unkorrelierte Rauschen stellt sich bei $T_{60} = 0\,\mathrm{s}$ der theoretisch maximale SNR-Gewinn von $10 \cdot \log_{10}(M) \simeq 7\,\mathrm{dB}$ ein, der mit steigendem Nachhall leicht abfällt. Für die gerichtete Störung ist für geringe Nachhallzeiten eine sehr hohe Unterdrückung des Störgeräusches möglich, da an der Stelle $\theta = \theta_n$ das *Beampattern* ein deutliches Minimum ausbildet (siehe Bild 4.5). Der sich einstellende SNR-Gewinn ist dabei von mehreren Faktoren abhängig wie der geo-

[11] Um sicherzustellen, dass die KLDS-Matrix der Störung invertierbar ist, wurde ein Regulisierungsterm von -30 dB eingefügt: $\mathbf{\Phi_{NN}}(\Omega) \leftarrow \mathbf{\Phi_{NN}}(\Omega) + 0,001 \cdot \sigma_N^2(\Omega) \cdot \mathbf{I}$, mit $\sigma_N^2(\Omega) = \mathrm{Spur}\{\mathbf{\Phi_{NN}}(\Omega)\}/\mathrm{M}$.

metrischen Anordnung, der Anzahl der Mikrophone und der spektralen Zusammensetzung der Störquelle. Wichtig an dieser Stelle ist lediglich der Vergleich zwischen den Verläufen von MVDR und GMVDR. Und dabei zeigt sich kein signifikanter Unterschied[12]

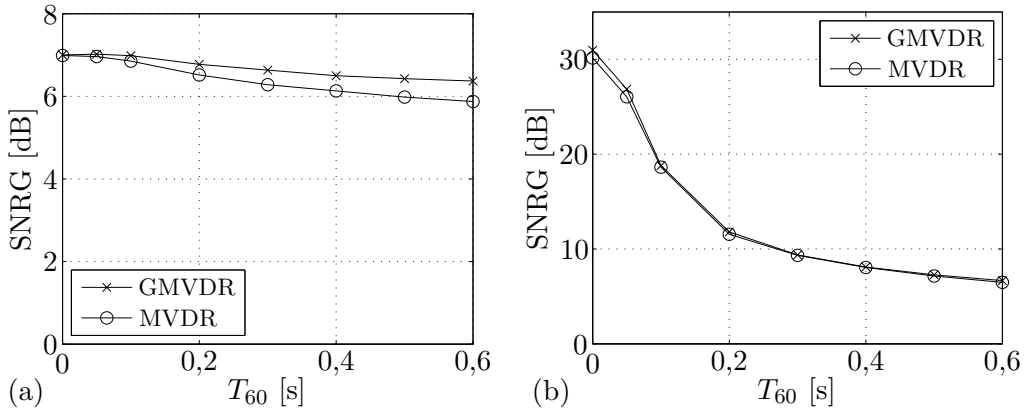

Bild 4.4: SNR-Gewinn für den MVDR und den verallgemeinerten MVDR *Beamformer*. (a) Störung besteht nur aus unkorreliertem Rauschen. (b) Gerichtetes Tiefpassrauschen als Störquelle.

Abgesehen von den geometrischen Verhältnissen und den spektralen Eigenschaften der Störung ist die Genauigkeit[13] der Schätzungen der Übertragungsfunktion $\mathbf{H}(\Omega)$ und der KLDS-Matrix $\mathbf{\Phi_{NN}}(\Omega)$ bzw. ihrer Inversen von entscheidender Bedeutung für die erzielbare Störgeräuschunterdrückung [Krü07]. Die in Bild 4.4 gezeigten Ergebnisse wurden mit einer Blocklänge von $L = 512$ berechnet. Dies führt jedoch bei höheren Nachhallzeiten aufgrund einer zu geringen Frequenzauflösung zu ungenauen Schätzungen von $\mathbf{\Phi_{NN}}(\Omega)$. Um aber eine annähernd korrekte Schätzung zu erhalten ist nach [JN87] eine Blocklänge von $L > f_{\mathrm{Ab}}/4 \cdot T_{60}$ notwendig. Wird diese nicht eingehalten, so ist mit Abstrichen in der resultierenden Störgeräuschreduktion zu rechnen. Dieser Zusammenhang kann durch folgende Betrachtungen veranschaulicht werden. Unter der Annahme einer korrekt geschätzten frequenzkontinuierlichen KLDS-Matrix ergibt sich diese für eine gerichtete Störung mit der Varianz $\sigma^2_{N,c}(\Omega)$ und der Übertragungsfunktion $\mathbf{A}(\Omega)$, sowie der Varianz $\sigma^2_{N,u}(\Omega)$ für das unkorrelierte Rauschen zu

$$\mathbf{\Phi_{NN}}(\Omega) = \sigma^2_{N,c}(\Omega)\mathbf{A}(\Omega)\mathbf{A}^H(\Omega) + \sigma^2_{N,u}(\Omega)\mathbf{I} \tag{4.58}$$

$$= \sum_{i=1}^{M} \lambda_i(\Omega)\mathbf{v}_i(\Omega)\mathbf{v}_i^H(\Omega). \tag{4.59}$$

In Gl. (4.59) ist mit $\lambda_i(\Omega)$ der $i-$te Eigenwert von $\mathbf{\Phi_{NN}}(\Omega)$ und mit $\mathbf{v}_i(\Omega)$ der zugehörige Eigenvektor bezeichnet. Die Eigenwerte sind reellwertig und seien hier, wie in den weiteren Kapiteln der Größe nach geordnet

$$\lambda_1(\Omega) \geq \lambda_2(\Omega) \geq \ldots \geq \lambda_M(\Omega) \geq 0. \tag{4.60}$$

Da jeder Vektor Eigenvektor einer Einheitsmatrix ist, gilt dies auch für den Vektor definiert

[12] Bei den hier gemachten Vergleichen zwischen den GMVDR und MVDR *Beamformern* soll nochmals darauf hingewiesen werden, dass die Filterkoeffizienten optimal mit den reinen Sprachdaten berechnet wurden.

[13] Die Genauigkeit der geschätzten Übertragungsfunktionen kann hier nicht explizit untersucht werden, da die zur Erzeugung der Sprachdaten verwendeten Impulsantworten nicht dirkt zu einem Vergleich zu verwenden sind. Diese sind deutlich länger und haben einen beliebigen Alpass-Anteil.

durch die gerichtete Störung

$$\mathbf{v}_1(\Omega) = \frac{\mathbf{A}(\Omega)}{||\mathbf{A}(\Omega)||}. \tag{4.61}$$

Dieser korrespondiert im Zusammenhang mit Gl. (4.58) zum größten Eigenwert

$$\lambda_1(\Omega) = \sigma^2_{N,c}(\Omega) \cdot ||\mathbf{A}(\Omega)||^2 + \sigma^2_{N,u}(\Omega) \tag{4.62}$$

und für alle anderen Eigenwerte gilt

$$\lambda_i(\Omega) = \sigma^2_{N,u}(\Omega),\ i = 2, \ldots, M. \tag{4.63}$$

Für die Inverse gilt folgende allgemeine Form

$$\mathbf{\Phi}^{-1}_{\mathbf{NN}}(\Omega) = \sum_{i=1}^{M} \frac{1}{\lambda_i}(\Omega)\mathbf{v}_i(\Omega)\mathbf{v}_i^H(\Omega). \tag{4.64}$$

Nun wird ein Eingangsvektor $\mathbf{X}(\Omega) = N_c(\Omega) \cdot \mathbf{A}(\Omega)$, welcher durch die gerichtete Störung hervorgerufen wird, angenommen und die Auswirkung dessen Filterung mit Gl. (4.28) untersucht. Für die Bildung des Minimums in der räumlichen Übertragungsfunktion und somit der Unterdrückung von Störgeräuschen der gerichteten Quelle ist bei der Anwendung von Gl. (4.28) im Wesentlichen die Rechtsmultiplikation von $\mathbf{\Phi}^{-1}_{\mathbf{NN}}(\Omega)$ mit $\mathbf{X}(\Omega)$ verantwortlich

$$\mathbf{\Phi}^{-1}_{\mathbf{NN}}(\Omega)\mathbf{X}(\Omega) = \left[\frac{\mathbf{A}(\Omega)\mathbf{A}^H(\Omega)}{||\mathbf{A}(\Omega)||^2\lambda_1(\Omega)} + \sum_{i=2}^{M} \frac{\mathbf{v}_i(\Omega)\mathbf{v}_i^H(\Omega)}{\lambda_i(\Omega)}\right] N_c(\Omega)\mathbf{A}(\Omega) \tag{4.65}$$

$$\approx \mathbf{0}, \quad \text{für } \sigma^2_{N,c} \gg \sigma^2_{N,u} \Leftrightarrow \lambda_1(\Omega) \gg \lambda_i(\Omega), i > 1. \tag{4.66}$$

Da die weiteren Eigenvektoren $\mathbf{v}_i(\Omega),\ i = 2, \ldots, M$ orthogonal zu $\mathbf{v}_1(\Omega)$ sind, ergibt sich also näherungsweise der Nullvektor für ein sehr kleines Verhältnis $\sigma^2_{N,u}(\Omega)/\sigma^2_{N,c}(\Omega)$, welches für ein letztes Experiment bezeichnet werden soll mit

$$\eta := 10 \log_{10} \frac{\sigma^2_{N,u}}{\sigma^2_{N,c}} \text{dB}. \tag{4.67}$$

Das Verhältnis η in Gl. (4.67) ist jedoch nicht mehr frequenzabhängig, sondern soll unter Berücksichtigung aller Frequenzkomponenten im Zeitbereich ermittelt werden. In Bild 4.5 (a) ist der SNR-Gewinn für ein variierendes η im Bereich zwischen -50 dB und 20 dB dargestellt, wobei wieder das *Szenario-2* zugrunde liegt und eine Nachhallzeit von $T_{60} = 0{,}05\,\text{s}$ gewählt wurde. Es ist deutlich zu erkennen, dass für ein steigendes η der SNR-Gewinn sinkt und gegen den Wert $10 \cdot \log_{10}(M) \simeq 7\,\text{dB}$ läuft. Das Bild 4.5 (b) verdeutlicht den Effekt der räumlichen Filterung. Zu sehen ist das *Beampattern* ausgewertet für eine Frequenz von ca. 1 kHz für unterschiedliche Verhältnisse der Varianzen Gl. (4.67). Das räumliche Minimum ist umso ausgeprägter, je größer die Varianz der korrelierten Störung im Vergleich zum unkorrelierten Rauschen ist. Für den Grenzwert $\eta \to -\infty$ hat die Matrix $\mathbf{\Phi}^{-1}_{\mathbf{NN}}(\Omega)$ den Rang eins, also alle Eigenwerte $\lambda_i(\Omega),\ i = 2, \ldots, M$ verschwinden und das *Beampattern* an der Stelle der Störquelle geht gegen $-\infty$.

Die explizite Betrachtung von Gl. (4.65) unter der Berücksichtigung von Gl. (4.58) zeigt die Degradation der Störgeräuschunterdrückung von räumlich korrelierten Störschallfeldern mit steigender Varianz unkorrelierter Störungen. Das Verhältnis $\sigma^2_{N,u}/\sigma^2_{N,c}$ wird in der Praxis beeinflusst durch ein variierendes $\sigma^2_{N,c}$ bei gleichbleibendem $\sigma^2_{N,u}$ (hervorgerufen durch z. B.

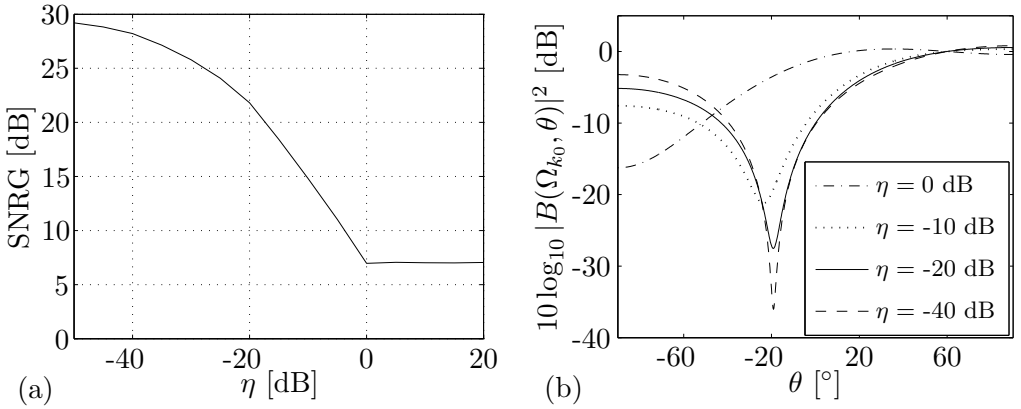

Bild 4.5: Analyse des variierenden Verhältnisses η der Varianzen der gerichteten und unkorrelierten Störung bei $T_{60} = 0,05$ s.(a) SNR-Gewinn und (b) *Beampattern* für eine Frequenz von ca. 1 kHz.

Mikrophonrauschen) und durch steigende Nachhallzeiten, so dass der diffuse Anteil des Störschallfeldes für höhere Frequenzen einen Beitrag zur unkorrelierten Störung leistet. Weiterhin ist natürlich eine möglichst genaue Schätzung von $\mathbf{\Phi_{NN}}(\Omega)$ bzw. ihrer Inversen notwendig. Mit steigender Nachhallzeit schleicht sich hier jedoch aufgrund zu kurzer Analysefenster ein systematischer Fehler ein, weshalb $\mathbf{v}_1(\Omega) = \mathbf{A}(\Omega)/||\mathbf{A}(\Omega)||$ für Gl. (4.65) immer ungenauer geschätzt wird.

4.6 Zusammenfassung und Diskussion

Für die mehrkanalige Geräuschreduktion mittels *Beamforming* wurden in diesem Kapitel statistisch optimale Filterkoeffizienten im Frequenzbereich hergeleitet. Dabei kam eine konsequente Schreibweise der verallgemeinerten Zusammenhänge zum Tragen, also die Verwendung der Raumübertragungsfunktion $\mathbf{H}(\Omega)$ zwischen dem Sprecher und der Mikrophongruppe, anstatt der Vereinfachung durch den *Steering Vector*. Die hier gezeigten unterschiedlichen Ansätze Max-SNR-Kriterium, Minimierung der Varianz, Maximierung der Plausibilität und Minimierung des kleinsten mittleren quadratischen Fehlers führen alle zu den gleichen optimalen Filterkoeffizienten bezüglich der räumlichen Selektivität und unterscheiden sich gerade in einem skalaren Faktor, welcher als spektrale, einkanalige Nachfilterung betrachtet werden kann. Wesentliche Unterschiede ergeben sich letztlich bei der Wahl des Adaptionsverfahrens[14] und der konkreten Implementierung.

Das Max-SNR-Kriterium unterscheidet sich jedoch von den anderen Verfahren dadurch, dass ein verallgemeinertes Eigenwertproblem gelöst werden kann und hierfür keinerlei Wissen über die Sprecherposition und die *Array*-Geometrie notwendig ist, weshalb es auch als "blindes" Verfahren bezeichnet werden kann. Die resultierenden Filterkoeffizienten beinhalten implizit eine Schätzung der Raumübertragungsfunktion. Diese Eigenschaft bringt jedoch auch einen entscheidenden Nachteil mit sich: da für ein breitbandiges Sprachsignal[15] die

[14] Die in diesem Kapitel aufgezeigten Lösungen für die optimalen Filterkoeffizienten ergeben sich bei einer entsprechenden Implementierung nach der Konvergenz der Koeffizienten. Dies kann mit unterschiedlichen Adaptionsverfahren erreicht werden (siehe z. B. Abschnitt 5).

[15] In der Antennentechnik werden aufgrund der schmalbandigen Signale Strukturen mittels Eigenwertzerlegung bedeutend häufiger diskutiert als bei der breitbandigen Sprachsignalverarbeitung (siehe z. B. [HBD00, Has02, EK03, YOZC04]).

Eigenwert-Dekomposition für jede Frequenz unabhängig voneinander erfolgt, können gravierende Sprachverzerrungen auftreten. Hier kann eine einkanalige Nachfilterung deutliche Abhilfe schaffen, welche einen Zusammenhang zu dem GMVDR-Verfahren herstellen soll. Auf Möglichkeiten der Realisierung eines solchen *Post Filters* wird in Kapitel 6 eingegangen.

Da die explizite Schätzung der Raumübertragungsfunktion bzw. einzelner Ausbreitungspfade in einer stark verhallten Umgebung sehr schwierig ist, werden solche Ansätze zur konstruktiven Nutzung der Mehrwegeausbreitung nur vereinzelt in der Literatur diskutiert. Eine frühe Arbeit, welche sich mit der Schätzung ausgeprägter Reflektionen beschäftigt, ist in [FSJ93] zu finden und führte zum so genannten *Matched Filter Array* [JF96]. Diese eher theoretisch angesiedelten Simulationen (*Array* mit 200 Sensoren) wurde in [RRFM98] weiter untersucht. In [AG97] fand der *Matched-Filter*-Ansatz eine Anwendung in einer GSC-Struktur für einen PC-Arbeitsplatz und einer expliziten Berücksichtigung von *Double-Talk*-Situationen in [AG96].

In [NNS01] ist ein Verfahren beschrieben, um multiple *Beamformer*, ausgerichtet auf den direkten Pfad und frühe Reflexionen, zu kombinieren. Ein ähnlicher Ansatz findet in [KHJ06] Anwendung. Hier wird wieder eine explizite Schätzung mehrerer Ausbreitungspfade verwendet, um sequentiell kaskadierte MVDR *Beamformer* zu adaptieren.

Weitere erfolgreiche Ansätze zur Ausnutzung der Mehrwegeausbreitung sind im Zusammenhang mit einer GSC-Struktur zu finden (siehe Kapitel 8). In [HSH99, HS01] werden adaptive Filter verwendet, um das verhallte Nutzsignal aus den Eingangssignalen herauszufiltern (*Blocking Matrix*) und so Störreferenzsignale zu erzeugen, die einen möglichst geringen Anteil des Sprachsignals enthalten. Dieser Ansatz findet in [HK01] eine effiziente Realisierung im Frequenzbereich und ist in [HK03] mit einer mehrkanaligen Echokompensation kombiniert. Eine zusätzliche Erweiterung zur Robustheitssteigerung bei impulsartigen Störungen in *Double-Talk*-Situationen wird in [HBNK07] beschrieben. In [GBW01] wird ein Verfahren vorgeschlagen um das Verhältnis der Übertragungsfunktionen (engl. *Transfer Function Ratio*) durch Ausnutzung der relativen Stationarität der Übertragungsfunktionen im Vergleich zu dem Nutzsignal zu schätzen und so ebenfalls Störreferenzsignale zu erzeugen. Dieser Ansatz ist in [GC04] mit einem *Post Filter* zur weiteren Störgeräuschreduktion kombiniert.

Kapitel 5

Adaptive Lösung des Eigenwertproblems

Die Berechnung der optimalen Filterkoeffizienten nach dem Max-SNR-Kriterium im laufenden Betrieb erfordert eine iterative Lösung des Eigenwertproblems Gl. (4.15) um eine adaptive Nachführung der Filterkoeffizienten zu gewährleisten. Grundvoraussetzung hierfür ist eine robuste Sprache/Pause-Detektion (siehe Anhang D), um einerseits das Kreuzleistungsdichtespektrum des Störschallfeldes und andererseits das Kreuzleistungsdichtespektrum aus der Überlagerung von Stör- und Nutzsignal zu schätzen.

Im Folgenden soll zunächst eine Untersuchung des speziellen Eigenwertproblems und anschließend des allgemeinen Eigenwertproblems erfolgen. Dafür werden Methoden vorgestellt und analysiert, die einerseits über Fixpunktverfahren und andererseits über Gradientenverfahren einen Eigenvektor korrespondierend zum größten Eigenwert einer Matrix iterativ bestimmen. Weiterhin muss die Unterscheidung gemacht werden, ob die Statistik der Eingangsdaten sich nicht mehr ändert und davon ausgegangen wird, dass die entsprechenden Matrizen deterministisch vorliegen. Oder, wie im Falle des akustischen *Beamformings*, die statistischen Eigenschaften der Signale sich über die Zeit sehr wohl ändern, weshalb der Übergang zu stochastischen Iterationsvorschriften gemacht werden muss. Experimentelle Untersuchungen bezüglich des Konvergenzverhaltens von Verfahren aus der Literatur und eigenentwickelten Verfahren zur iterativen Bestimmung des gesuchten Eigenvektors sollen hier durchgeführt werden.

Da die iterative Schätzung des gesuchten Eigenvektors für den frequenzdiskreten Fall umgesetzt werden soll, und diese für jede Frequenzkomponente unabhängig voneinander durchzuführen ist, wird in diesem Kapitel auf einen frequenzabhängigen Parameter verzichtet. Dies erhöht die Lesbarkeit, insbesondere, da ein zusätzlicher Index für den Iterationsschritt eingeführt werden muss.

5.1 Spezielles Eigenwertproblem

Die grundlegende Thematik dieses Abschnitts ist mit der Formulierung des speziellen Eigenwertproblems

$$\mathbf{\Phi_{XX}}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad 1 \leq i \leq M \tag{5.1}$$

gegeben. Es sei angemerkt, dass die $\mathbf{v}_i$ in Gl. (5.1) nicht eindeutig bestimmt sind, da die Eigenwertgleichung ebenfalls für alle Vektoren $\zeta\mathbf{v}_i$ mit dem komplexwertigen Skalar ζ gilt. Außerdem existieren für beliebige Matrizen $\mathbf{\Phi_{XX}}$ der Dimension $M \times M$ nicht immer M

unabhängige Eigenvektoren. Daher wird hier und im Folgenden immer davon ausgegangen, dass die Eigenvektoren auf die Einheitslänge normiert sein sollen

$$||\mathbf{v}_i|| = 1, \quad \forall i. \tag{5.2}$$

Obschon Lösungsvorschläge für die Problemstellung Gl. (5.1) seit über 160 Jahren[1] in der Literatur diskutiert werden, ist nach wie vor die iterative Lösung des Eigenwertproblems Gegenstand aktueller Forschungsarbeiten aufgrund der hohen Relevanz im Bereich der numerischen, linearen Algebra, siehe z. B. [GV00, CA03]. In dieser Arbeit ist von einer positiv definiten, hermiteschen Matrix $\mathbf{\Phi_{XX}}$ der Dimension $M \times M$ auszugehen, so dass die M Eigenwerte λ_i positiv und reell sind. Diese seien der Größe nach angeordnet

$$\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_M \geq 0. \tag{5.3}$$

Weiterhin ist im Rahmen dieser Arbeit nur ein Eigenvektor korrespondierend zum größten Eigenwert $\lambda^{(\max)} = \lambda_1$ zu bestimmen (engl. *Principal Component Analysis*, PCA). Dieser trägt gemäß der Nummerierung in Gl. (5.3) den Index Eins ($\mathbf{v}_1$) und entspricht gerade dem gesuchten Filterkoeffizientenvektor $\mathbf{F}$. Diese Definition entspricht der Annahme, dass in der allgemeinen Betrachtung Gl. (4.15) die KLDS-Matrix der Störung nicht berücksichtigt wird

$$\mathbf{\Phi_{XX}F} = \lambda^{(\max)}\mathbf{F}. \tag{5.4}$$

Für das *Beamforming* ist diese Formulierung äquivalent zur Ausrichtung der Hauptkeule des *Beampatterns* in Richtung der dominanten Quelle. Für die Betrachtung der drei möglichen Arten von Störschallfeldern hat dies folgende Bedeutung:

- Da die unkorrelierte Störung keinerlei Einfluss auf die "Richtung" von $\mathbf{v}_1$ hat, sondern lediglich auf dessen Skalierung, ergibt sich an dieser Stelle keinerlei Informationsverlust.
- Im Falle des diffusen Störschallfeldes werden die frequenzabhängigen Hauptkeulen ebenfalls korrekt auf den Zielsprecher ausgerichtet. Jedoch ergibt sich hier unter Vernachlässigung des Kohärenzterms Gl. (2.20) ein Verlust bezüglich des maximal erzielbaren SNR-Gewinns aufgrund der reduzierten Direktivität. Da jedoch grundsätzlich das Nutzsignal in den einzelnen Signalpfaden nach der Filterung mit den *Beamformer*-Koeffizienten kohärent vorliegt, kann eine Nachfilterung ähnlich zum superdirektiven *Beamforming* zur Steigerung der Direktivität vorgenommen werden.
- Ist im Raum jedoch eine starke, gerichtete Störung vorhanden, so wird das frequenzabhängige *Beampattern*, gegeben durch die Lösung von Gl. (5.4), sich entweder auf den Sprecher oder auf die Störung, bzw. einer Mischung aus beiden, ausrichten. Für diesen Fall ist die Lösung des allgemeinen Eigenwertproblems Gl. (4.15) unerlässlich. Daher kann eine PCA-Adaption nur eingesetzt werden wenn keine starken Störer vorhanden sind. Dies ist über eine SNR abhängige Steuerung sicherzustellen.

Dies bedeutet also, dass im Falle nicht vorhandener gerichteter Störquellen, oder wenn diese zumindest im Vergleich zum Sprachsignal nur eine sehr geringe Leistung emittieren, durch die Lösung des speziellen Eigenwertproblems die Filterkoeffizienten eine Matched Filterung vornehmen und somit quasi ein "selbstjustierender" DSB realisiert werden kann. Dessen Direktivität kann durch eine geeignete Nachfilterung noch erhöht werden.

[1] Im Jahre 1846 erschien bereits eine wichtige Arbeit von Jacobi [Jac46] zur Lösung des Eigenwertproblems. Da die Matrixnotation damals noch unbekannt war, formulierte er das Problem allerdings durch elementweise Betrachtung von Systemgleichungen.

5.1.1 Potenzmethode

Zunächst soll davon ausgegangen werden, dass $\mathbf{\Phi_{XX}}$ aus der blockweisen Verarbeitung der eingehenden Mikrophonsignale $\mathbf{X}_m$ mit dem Blockindex m bestimmt worden ist

$$\mathbf{\Phi_{XX}} = \sum_{i=1}^{M} \lambda_i \mathbf{v}_i \mathbf{v}_i^H = \lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \mathbf{X}_m \mathbf{X}_m^H \tag{5.5}$$

und somit deterministische Methoden verwendet werden können. Als Motivation für die Potenzmethode kann nun folgende Vorgehensweise gesehen werden. Für den gesuchten Eigenvektor gilt unter Berücksichtigung von Gl. (5.2)

$$\mathbf{\Phi_{XX}}\mathbf{v}_1 = \lambda_1 \mathbf{v}_1 \tag{5.6}$$

$$\frac{\mathbf{\Phi_{XX}}\mathbf{v}_1}{||\mathbf{\Phi_{XX}}\mathbf{v}_1||} = \mathbf{v}_1. \tag{5.7}$$

Mit der Einführung des Iterationszählers κ, ergibt sich das einfache Iterationsverfahren der Potenzmethode[2] zu

$$\hat{\mathbf{v}}_{1,\kappa} = \frac{\mathbf{\Phi_{XX}}\hat{\mathbf{v}}_{1,\kappa-1}}{||\mathbf{\Phi_{XX}}\hat{\mathbf{v}}_{1,\kappa-1}||}, \quad \kappa = 1, 2, \ldots \tag{5.8}$$

mit dem Startvektor[3]

$$\hat{\mathbf{v}}_{1,0} = \sum_{i=1}^{M} c_i \mathbf{v}_i, \quad c_i \in \mathbb{C}, c_1 \neq 0. \tag{5.9}$$

Das Konvergenzverhalten kann anschaulich an der Folge $\mathbf{\Phi_{XX}}\hat{\mathbf{v}}_{1,0}, \mathbf{\Phi_{XX}}(\mathbf{\Phi_{XX}}\hat{\mathbf{v}}_{1,0}), \ldots$ betrachtet werden. Es gilt für den κ-ten Schritt

$$\begin{aligned} \mathbf{\Phi}_{\mathbf{XX}}^{\kappa}\hat{\mathbf{v}}_{1,0} &= \sum_{i=1}^{M} c_i \lambda_i^{\kappa} \mathbf{v}_i \\ &= \lambda_1^{\kappa} \left[c_1 \mathbf{v}_1 + \sum_{i=2}^{M} c_i \left(\frac{\lambda_i}{\lambda_1} \right)^{\kappa} \mathbf{v}_i \right]. \end{aligned} \tag{5.10}$$

Mit der Annahme Gl. (5.3) erkennt man, dass der rechte Term in Gl. (5.10) für ein steigendes κ verschwindet und somit nur noch eine Komponente in die Richtung $\mathbf{v}_1$ übrig bleibt. Die Folge $\{\hat{\mathbf{v}}_{1,\kappa}\}_{\kappa\in\mathbb{N}}$ in Gl. (5.8) konvergiert also linear gegen $c_1/|c_1|\mathbf{v}_1$ mit der Konvergenzrate λ_2/λ_1, da der Ausdruck $(\lambda_2/\lambda_1)^{\kappa}$ in Gl. (5.10) am langsamsten gegen Null strebt. Es bleibt noch anzumerken, dass der Fehler zwischen zwei Iterationen von der Wahl der Startwerte c_i abhängt, wie an Gl. (5.10) ebenfalls abgelesen werden kann.

Anhand der vorhergehenden Betrachtungen liegt der wesentliche Nachteil der Potenzmethode klar auf der Hand. Liegen die Eigenwerte nahe beieinander, so konvergiert die Folge Gl. (5.8) nur sehr langsam. Abhilfe verschaffen hier zahlreiche Verfahren, welche in der Literatur der letzten Jahrzehnte zu finden sind. Diese sind jedoch bedeutend komplexer vom Rechenaufwand her oder gehen von bestimmten Annahmen an die Problemstellung aus. Liegt

[2] Housholder [Hou64] schreibt die erste Verwendung der Potenzmethode dem Mathematiker Müntz im Jahre 1913 zu. Zuvor wurde sie jedoch in [Bod56] dem Mathematiker von Mises und dessen Veröffentlichung [VMPG29] im Jahre 1929 zuerkannt. Daher wird die Potenzmethode auch als Vektoriteration nach von Mises bezeichnet.

[3] Es läßt sich keine Methode zur Bestimmung eines idealen Startvektors angeben. Als sinnvoll hat sich hier die Wahl eines rein reellen Vektors mit gleichen Einträgen für alle Elemente erwiesen.

z. B. eine gute Approximation der gesuchten Eigenwerte vor, so erreicht man mit der inversen Iteration nach Wielandt [Wie44] eine erhebliche Beschleunigung der Potenzmethode. Viele Methoden basieren auf Orthogonaltransformationsverfahren und beziehen die gesamte Iterationsfolge $\{\mathbf{\Phi}_{\mathbf{XX}}^{\kappa}\hat{\mathbf{v}}_{1,0}\}_{\kappa\in\mathbb{N}}$ in die Iteration ein, welche den so genannten Krylov Unterraum $\mathcal{K}$ bildet

$$\mathcal{K}_{\kappa}(\hat{\mathbf{v}}_{1,0};\mathbf{\Phi}_{\mathbf{XX}}) \equiv \text{span}\{\hat{\mathbf{v}}_{1,0},\mathbf{\Phi}_{\mathbf{XX}}\hat{\mathbf{v}}_{1,0},\ldots,\mathbf{\Phi}_{\mathbf{XX}}^{\kappa}\hat{\mathbf{v}}_{1,0}\}. \tag{5.11}$$

Mit $\text{span}\{\mathbf{u}_1,\ldots,\mathbf{u}_M\}$ ist ein Unterraum beschrieben, der durch die Vektoren $\mathbf{u}_1,\ldots,\mathbf{u}_M$ aufgespannt (engl. *span*) wird. Wichtige, grundlegende Verfahren sind hier das Lanczos-Verfahren [Lan50] sowie das Arnoldi-Verfahren [Arn51].

Für die Bestimmung aller Eigenwerte und Eigenvektoren des Eigenwertproblems kann z. B. der recht aufwendige QR-Algorithmus [Fra61] verwendet werden. Dabei wird in der Regel zuerst eine Hessenberg-Matrix (quadratische Matrix, deren Einträge unterhalb der ersten Nebendiagonalen gleich Null sind) berechnet und anschließend eine QR-Transformation[4] vorgenommen. Weitere Verfahren zur Eigenwertbestimmung können z. B. [GV00] entnommen werden.

Zusammenfassend lässt sich sagen, dass bei Matrizen geringer Ordnung und Interesse an lediglich eines Eigenvektors korrespondierend zum größten Eigenwert die Potenzmethode aufgrund der geringen Rechenkomplexität ein sehr effektives Verfahren darstellt. Wobei je nach Anwendung[5] auf die oben genannten Konvergenzeigenschaften zu achten ist. Für die Anwendung des akustischen *Beamformings* bedeutet dies:

- Für das Max-SNR-Kriterium ist nur ein Eigenvektor einer Matrix geringer Ordnung zu berechnen.
- Die Potenzmethode eignet sich auch bei schwach besetzten Matrizen.
- Der Rechenaufwand ist gering und eignet sich somit für Echtzeit-Anwendungen.
- In der Regel[6] gilt $\lambda_1 \gg \lambda_2$, wodurch eine Konvergenz sichergestellt ist.
- Die letztendliche Konvergenzgeschwindigkeit hängt maßgeblich von einer guten Schätzung der KLDS-Matrizen ab und weniger von der Konvergenzrate der Potenzmethode.

Nun soll der stochastische Fall betrachtet werden, für den statt der Matrix $\mathbf{\Phi}_{\mathbf{XX}}$ nur eine stochastische Schätzung $\hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}$ zum Iterationszeitpunkt κ vorliegt. Hierbei werden alle eingehenden Daten bis zum Iterationszeitpunkt für die Schätzung $\hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}$ verwendet. In der Regel ist dabei der Blockindex m gleichbedeutend mit dem Iterationsindex κ, so dass zwischen drei Möglichkeiten der Zeitreihenanalyse unterschieden werden kann: der Gleichmäßigen Gewichtung (GG), der Exponentiellen Glättung (EG) und der Instantanen Schätzung (IS).

Gleichmäßige Gewichtung (GG) Bei der gleichmäßigen Gewichtung bzw. dem gleitenden Mittelwert (engl. *Moving Average*) tragen alle eingehenden Daten innerhalb eines gewis-

[4] Wenn $\mathbf{A}$ eine gegebene Matrix mit linear unabhängigen Spalten ist, so gibt es eine Matrix $\mathbf{Q}$ mit orthogonalen Spalten und eine obere Dreiecksmatrix $\mathbf{R}$, so dass $\mathbf{A} = \mathbf{QR}$ gilt.

[5] Google benutzt z. B. die Potenzmethode zur Bewertung der relativen Wichtigkeit eines Links (*PageRank*).

[6] Für den Fall eines Ein-Sprecher-Szenarios gilt $\lambda_1 \gg \lambda_2$ (siehe alternativ blinde Quellentrennung [TV07]).

sen Zeitfensters N gleichstark zur rekursiven Schätzung bei

$$\hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},\kappa} = \begin{cases} \frac{\kappa-1}{\kappa}\hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},\kappa-1} + \frac{1}{\kappa}\mathbf{X}_\kappa\mathbf{X}^H_\kappa & \text{falls } 1 \le \kappa \le N, \\ \hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},\kappa-1} + \frac{1}{N}\left(\mathbf{X}_\kappa\mathbf{X}^H_\kappa - \mathbf{X}_{\kappa-N}\mathbf{X}^H_{\kappa-N}\right) & \text{sonst.} \end{cases} \tag{5.12}$$

Wählt man $N \to \infty$ so würde für alle Zeiten die Gesamtheit der Eingangsdaten gleichgewichtet berücksichtigt werden. Einerseits bedeutet dies, dass eine gute Approximation im Sinne von Gl. (5.5) anfällt, aber andererseits man Änderungen in der Statistik (z. B. Sprecherbewegungen) für große κ nicht erfassen würde. Dennoch wird diese Variante der konsistenteren Notation wegen mit der gleichmäßigen Gewichtung assoziiert.

Exponentielle Glättung (EG) Die exponentielle Glättung versieht Daten mit abnehmender Aktualität mit einem geringer werdenden Gewicht

$$\hat{\mathbf{\Phi}}^{(\mathrm{EG})}_{\mathbf{XX},\kappa} = \alpha\hat{\mathbf{\Phi}}^{(\mathrm{EG})}_{\mathbf{XX},\kappa-1} + (1-\alpha)\mathbf{X}_\kappa\mathbf{X}^H_\kappa, \quad 0 < \alpha < 1, \tag{5.13}$$

wobei die Glättungskonstante α nahe bei Eins liegt. Sie kann auch für eine gewünschte zeitliche Einwirktiefe τ_g der exponentiellen Glättung und gegebener Blocklänge L analytisch bestimmt werden mit

$$\alpha = 1 - \frac{L}{\tau_g \cdot f_{Ab}}. \tag{5.14}$$

Instantaner Schätzer (IS) Wird lediglich der aktuelle Eingangsblock verwendet, so liegt eine instantane Schätzung vor

$$\hat{\mathbf{\Phi}}^{(\mathrm{IS})}_{\mathbf{XX},\kappa} = \mathbf{X}_\kappa\mathbf{X}^H_\kappa. \tag{5.15}$$

Solch ein Vorgehen weist natürlich eine hohe Varianz der Schätzung auf, ermöglicht aber auch ein schnelles Reagieren auf eine sich ändernde Statistik der Eingangsdaten. In der Regel wird eine instantane Schätzung im Zusammenhang mit einer weiteren Mittelung oder mit Schrittweite-Verfahren verwendet.

Für die Iteration mittels der Potenzmethode bedeutet der stochastische Ansatz eine wechselseitige Aktualisierung von zuerst $\hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}$ und danach $\hat{\mathbf{v}}_{1,\kappa}$ aus Gl. (5.8). In [Kar84] ist diese wechselseitige Iteration mit Gl. (5.12) und $N \to \infty$ bereits explizit beschrieben. Hier sollen nun zwei Algorithmen für die stochastische Potenzmethode angegeben werden; zur Lösung des speziellen Eigenwertproblems mittels Potenzmethode und gleichmäßiger Gewichtung (S-PM-GG):

Algorithmus 1 (S-PM-GG) *Wähle die Fenstergröße N und einen Startvektor $\hat{\mathbf{v}}_{1,0} \in \mathbb{C}^M$. Berechne für $\kappa = 1, 2, \ldots$*

$$\begin{aligned} \hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},\kappa} &:= \begin{cases} \frac{\kappa-1}{\kappa}\hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},\kappa-1} + \frac{1}{\kappa}\mathbf{X}_\kappa\mathbf{X}^H_\kappa & \textit{falls } 1 \le \kappa \le N, \\ \hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},\kappa-1} + \frac{1}{N}\left(\mathbf{X}_\kappa\mathbf{X}^H_\kappa - \mathbf{X}_{\kappa-N}\mathbf{X}^H_{\kappa-N}\right) & \textit{sonst} \end{cases} \\ \mathbf{a} &:= \hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},\kappa}\hat{\mathbf{v}}_{1,\kappa-1} \\ \hat{\mathbf{v}}_{1,\kappa} &:= \frac{\mathbf{a}}{||\mathbf{a}||} \end{aligned}$$

sowie für das spezielle Eigenwertproblem mittels Potenzmethode und exponentieller Gewichtung (S-PM-EG):

Algorithmus 2 (S-PM-EG) *Wähle eine Glättungskonstante α und einen Startvektor $\hat{\mathbf{v}}_{1,0} \in \mathbb{C}^M$. Berechne für $\kappa = 1, 2, \ldots$*

$$\begin{aligned} \hat{\mathbf{\Phi}}^{(\mathrm{EG})}_{\mathbf{XX},\kappa} &:= \alpha \hat{\mathbf{\Phi}}^{(\mathrm{EG})}_{\mathbf{XX},\kappa-1} + (1-\alpha)\mathbf{X}_\kappa \mathbf{X}_\kappa^H \\ \mathbf{a} &:= \hat{\mathbf{\Phi}}^{(\mathrm{EG})}_{\mathbf{XX},\kappa} \hat{\mathbf{v}}_{1,\kappa-1} \\ \hat{\mathbf{v}}_{1,\kappa} &:= \frac{\mathbf{a}}{||\mathbf{a}||} \end{aligned}$$

5.1.2 Projektionsapproximation

Eine Reduzierung des Rechenaufwands der Potenzmethode für den stochastischen Fall mit gleichmäßiger Gewichtung zur Bestimmung von $\hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}$ ist in [RP02, RPW04] erläutert. Dabei wird entsprechend der Potenzmethode von folgendem Ausdruck ausgegangen

$$\hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},\kappa} \hat{\mathbf{v}}_{1,\kappa-1} = \frac{1}{\kappa} \sum_{m=1}^{\kappa} \mathbf{X}_m \mathbf{X}_m^H \hat{\mathbf{v}}_{1,\kappa-1}. \tag{5.16}$$

Die Projektion von $\mathbf{X}_m$ auf $\hat{\mathbf{v}}_{1,\kappa-1}$ in Gl. (5.16) wird dann wie folgt angenähert:

$$\mathbf{X}_m^H \hat{\mathbf{v}}_{1,\kappa-1} \approx \mathbf{X}_m^H \hat{\mathbf{v}}_{1,m-1} \qquad \forall m, \kappa \quad 1 \le m \le \kappa. \tag{5.17}$$

Die rechte Seite von Gl. (5.17) entspricht gerade der Filterung der Eingangsdaten für den Block m, also der Definition $Y_m^* := \mathbf{X}_m^H \hat{\mathbf{v}}_{1,m-1}$. Verwendet man nun diese Definition mit der Approximation Gl. (5.17) in Gl. (5.16) und setzt zusätzlich noch Gl. (5.12) ein ergibt sich der gleichgewichtete Projektionsvektor

$$\mathbf{p}_\kappa^{(\mathrm{GG})} = \frac{\kappa - 1}{\kappa} \mathbf{p}_{\kappa-1}^{(\mathrm{GG})} + \frac{1}{\kappa} \mathbf{X}_\kappa Y_\kappa^*, \tag{5.18}$$

bzw. mit Gl. (5.13) der exponentiell gewichtete Projektionsvektor

$$\mathbf{p}_\kappa^{(\mathrm{EG})} = \alpha \mathbf{p}_{\kappa-1}^{(\mathrm{EG})} + (1-\alpha) \mathbf{X}_\kappa Y_\kappa^*. \tag{5.19}$$

Die Iterationsvorschrift der stochastischen Potenzmethode mit Projektionsapproximation[7] besteht dann natürlich noch aus der Normierung

$$\hat{\mathbf{v}}_{1,\kappa} = \frac{\mathbf{p}_\kappa}{||\mathbf{p}_\kappa||}, \tag{5.20}$$

wobei der hochgestellte Index für die Bezeichnung der Glättung in Gl. (5.20) nicht explizit aufgeführt ist. Es bleibt also festzuhalten, dass durch die Projektionsapproximation eine Komplexitätsreduktion von der Ordnung $\mathcal{O}(M^2)$ hin zu $\mathcal{O}(M)$ vorgenommen wurde.

[7] Projektionsmethoden, die eine Näherung für die Schätzung von Eigenräumen von Matrizen bilden, sind auch unter dem Begriff *Projection Approximation Subspace Tracking* (PAST) Verfahren bekannt, auch wenn diese Begrifflichkeit in [RP02, RPW04] nicht explizit fällt.

5.1.3 Gradientenverfahren

Setzt man für die Glättungskonstante α in Gl. (5.19) einen Wert sehr nahe 1 an und bezeichnet $1-\alpha$ als Schrittweite μ, so kann Gl. (5.19) wiederum approximiert werden durch

$$\mathbf{p}_\kappa = \mathbf{p}_{\kappa-1} + \mu \mathbf{X}_\kappa (\mathbf{X}_\kappa^H \mathbf{p}_{\kappa-1}). \tag{5.21}$$

Es ergibt sich damit ein Zusammenhang zu dem so genannten Hebbschen Postulat[8] des Physiologen Donald Hebb von 1949. Darin beschreibt er prinzipiell die Regel Gl. (5.21) mit dem Begriff Effizienz, welche in seinem Fall die synaptische Veränderung zwischen Nervenzellen meint. Die Interpretation von Gl. (5.21) ist nun derart, dass es sich um ein Gradientenanstiegsverfahren handelt, welches die Ausgangsenergie $|\tilde{Y}_\kappa|^2 := |\mathbf{p}_{\kappa-1}^H \mathbf{X}_\kappa|^2$ mit fortlaufender Iteration prinzipiell unendlich stark anwachsen lässt. Definiert man die Kostenfunktion

$$J(\mathbf{p}_{\kappa-1}) = \mathbf{p}_{\kappa-1}^H \hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}^{(\mathrm{IS})} \mathbf{p}_{\kappa-1}, \tag{5.22}$$

welche durch geeignete Wahl von $\mathbf{p}_{\kappa-1}$ zu maximieren ist, so kann mit

$$\nabla_{\mathbf{p}} J(\mathbf{p}_{\kappa-1}) = 2\hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}^{(\mathrm{IS})} \mathbf{p}_{\kappa-1} \tag{5.23}$$

an der allgemeinen Lernregel für das Gradientenanstiegsverfahren

$$\mathbf{p}_\kappa = \mathbf{p}_{\kappa-1} + \frac{\mu}{2} \nabla_{\mathbf{p}} J(\mathbf{p}_{\kappa-1}) \tag{5.24}$$

und Gl. (5.15) die Gleichheit von Gl. (5.21) und Gl. (5.24) eerkannt werden. Das Problem des unbegrenzten Anwachsens von $\mathbf{p}_\kappa$ ist in [Ama77] intuitiv mit einer expliziten Normierung des Koeffizientenvektors wie folgt gelöst:

$$Y_\kappa := \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{X}_\kappa \quad \rightarrow \quad \mathbf{p}_{1,\kappa} = \mathbf{p}_{1,\kappa-1} + \mu \mathbf{X}_\kappa Y_\kappa^* \quad \rightarrow \quad \hat{\mathbf{v}}_{1,\kappa} = \frac{\mathbf{p}_\kappa}{||\mathbf{p}_\kappa||}. \tag{5.25}$$

Interessanterweise ist der explizite Normierungsschritt in Gl. (5.25) in [Oja82] implizit in die Herleitung der Gradientenanstiegsmethode eingebunden. Diese Vorschrift wird wegen [Oja82] auch synonym Ojas-Regel genannt. Für die folgende Herleitung des Algorithmus soll zunächst der deterministische Ansatz hergenommen werden. Dafür soll eine Maximierungsaufgabe mit Randbedingung formuliert werden:

$$\max_{\mathbf{v}^H} \mathbf{v}^H \mathbf{\Phi}_{\mathbf{XX}} \mathbf{v} \quad \text{unter der Randbed.} \quad \mathbf{v}^H \mathbf{v} = C^2, \quad C \in \mathbb{R}^+. \tag{5.26}$$

Die Norm der Filterkoeffizienten ist also durch den reellwertigen Parameter C festgelegt. Es soll nun eine reelle Kostenfunktion definiert werden, welche im Vergleich zu Gl. (5.22) die Randbedingung durch den reellwertigen Lagrange-Multiplikator β beinhaltet

$$J(\mathbf{v}) = \mathbf{v}^H \mathbf{\Phi}_{\mathbf{XX}} \mathbf{v} + \beta(\mathbf{v}^H \mathbf{v} - C^2). \tag{5.27}$$

Für den Gradienten von $J(\mathbf{v})$ bezüglich den gesuchten Koeffizienten $\mathbf{v}$ ergibt sich

$$\nabla_{\mathbf{v}} J = 2\mathbf{\Phi}_{\mathbf{XX}} \mathbf{v} + 2\beta \mathbf{v}, \tag{5.28}$$

[8] *"Wenn ein Axon der Zelle A nahe genug ist, um eine Zelle B zu erregen und wiederholt oder dauerhaft sich am Feuern beteiligt, geschieht ein Wachstumsprozess oder metabolische Änderung in einer oder beiden Zellen derart, dass A's Effizienz, als eine der auf B feuernde Zellen, anwächst."* (frei Übersetzt nach D. Hebb, 1949)

welcher zu Null zu setzen ist. Dadurch lässt sich schließlich mit Einhaltung der Nebenbedingung der gesuchte Faktor β berechnen

$$\beta = \frac{\mathbf{v}^H \mathbf{\Phi_{XX}} \mathbf{v}}{C^2}. \tag{5.29}$$

Die Iterationsgleichung für $\hat{\mathbf{v}}_{1,\kappa}$ mittels determinstischem Gradientenanstieg und Gl. (5.28) sowie Gl. (5.29) ist

$$\begin{aligned} \hat{\mathbf{v}}_{1,\kappa} &= \hat{\mathbf{v}}_{1,\kappa-1} + \frac{\mu}{2} \nabla_{\mathbf{v}} J \Big|_{\mathbf{v}=\hat{\mathbf{v}}_{1,\kappa-1}} & (5.30) \\ &= \hat{\mathbf{v}}_{1,\kappa-1} + \mu \left(\mathbf{\Phi_{XX}} \hat{\mathbf{v}}_{1,\kappa-1} - \frac{\hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{XX}} \hat{\mathbf{v}}_{1,\kappa-1}}{C^2} \hat{\mathbf{v}}_{1,\kappa-1} \right). & (5.31) \end{aligned}$$

In [OK85, CHY98] wurde gezeigt, dass Oja's Regel Gl. (5.31) gegen den gewünschten Eigenvektor konvergiert, und also der Fehlerterm in der Klammer für $\kappa \to \infty$ verschwindet. Setzt man nun die instantane Schätzung Gl. (5.15) in Gl. (5.31) ein, so ergibt sich mit $Y_\kappa = \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{X}_\kappa$ die stochastische Regel

$$\hat{\mathbf{v}}_{1,\kappa} = \hat{\mathbf{v}}_{1,\kappa-1} + \mu Y_\kappa^* \left(\mathbf{X}_\kappa - \frac{Y_\kappa}{C^2} \hat{\mathbf{v}}_{1,\kappa} \right), \tag{5.32}$$

welche nach wie vor eine der beliebtesten Iterationsregeln zur Schätzung des Eigenvektors $\mathbf{v}_1$ darstellt. Für Gl. (5.32) bleibt der Koeffizientenvektor im stationären Zustand ($\hat{\mathbf{v}}_{1,\kappa} = \hat{\mathbf{v}}_{1,\kappa-1}$) unter folgenden Bedingungen: (i) $\hat{\mathbf{v}}_{1,\kappa} = C\mathbf{v}_1$, (ii) $\hat{\mathbf{v}}_{1,\kappa}^H \hat{\mathbf{v}}_{1,\kappa} = C^2$ und (iii) $X_\kappa = c_{1,\kappa}\mathbf{v}_1$ mit $c_{1,\kappa} \in \mathbb{C}$. Da die Bedingung (iii) bei der Adaption sicherlich nicht für alle Eingangsdaten zutrifft, wird die Schätzung je nach Größe der Schrittweite μ um den gesuchten Eigenvektor herum schwanken. Es sei noch angemerkt, dass üblicherweise die Nebenbedingung (engl. *Constraint*) zu $C = 1$ gesetzt wird.

5.1.4 Neuartiges Gradientenverfahren

Die Herleitung eines neuen Gradientenverfahrens zur iterativen Bestimmung von $\mathbf{v}_1$ basiert ebenfalls auf der Maximierungsaufgabe Gl. (5.26), der Kostenfunktion Gl. (5.27) und der Gradientenanstieg-Methode Gl. (5.30). Das Verfahren wurde erstmals in [WHU05] präsentiert und für das akustische *Beamforming* eingesetzt. Der Lagrange-Multiplikator wird jedoch mittels der Bedingung

$$\hat{\mathbf{v}}_{1,\kappa}^H \hat{\mathbf{v}}_{1,\kappa} \stackrel{!}{=} C^2, \quad C \in \mathbb{R}^+. \tag{5.33}$$

berechnet, also der Einhaltung der Nebenbedingung im nächsten Iterationsschritt:

$$\begin{aligned} C^2 &= \left[\hat{\mathbf{v}}_{1,\kappa-1}^H + \mu \hat{\mathbf{v}}_{1,\kappa-1}^H \left(\mathbf{\Phi_{XX}} + \beta \mathbf{I} \right) \right] \left[\hat{\mathbf{v}}_{1,\kappa-1} + \mu \left(\mathbf{\Phi_{XX}} + \beta \mathbf{I} \right) \hat{\mathbf{v}}_{1,\kappa-1} \right] \\ &\approx \hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1} + 2\mu \hat{\mathbf{v}}_{1,\kappa-1}^H \left(\mathbf{\Phi_{XX}} + \beta \mathbf{I} \right) \hat{\mathbf{v}}_{1,\kappa-1}, \end{aligned} \tag{5.34}$$

wobei der Term mit μ^2 in der Approximation Gl. (5.34) vernachlässigt wurde (aufgrund von $\mu < 10^{-4}$). Man erhält schließlich für den Lagrange-Multiplikator

$$\beta \approx \frac{C^2 - \hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1} - 2\mu \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{XX}} \hat{\mathbf{v}}_{1,\kappa-1}}{2\mu \hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1}}. \tag{5.35}$$

Setzt man Gl. (5.35) in Gl. (5.28) ein und benutzt die Iterationsgleichung Gl. (5.30), ergibt sich nach einiger Rechnung

$$\hat{\mathbf{v}}_{1,\kappa} = \frac{C^2 + \hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1}}{2\hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1}} \hat{\mathbf{v}}_{1,\kappa-1} + \mu \left(\mathbf{\Phi}_{\mathbf{XX}} \hat{\mathbf{v}}_{1,\kappa-1} - \frac{\hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi}_{\mathbf{XX}} \hat{\mathbf{v}}_{1,\kappa-1}}{\hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1}} \hat{\mathbf{v}}_{1,\kappa-1} \right), \tag{5.36}$$

und äquivalent zu Gl. (5.32) kann auch hier mit $Y_\kappa = \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{X}_\kappa$ eine stochastische Adaptionsregel angegeben werden

$$\hat{\mathbf{v}}_{1,\kappa} = \frac{C^2 + \hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1}}{2\hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1}} \hat{\mathbf{v}}_{1,\kappa-1} + \mu Y_\kappa^* \left(\mathbf{X}_\kappa - \frac{Y_\kappa}{\hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1}} \hat{\mathbf{v}}_{1,\kappa-1} \right). \tag{5.37}$$

An dem neuen Algorithmus ist zuerst einmal eine Eigenschaft offensichtlich: Wird bei der Iteration die Nebenbedingung erfüllt, so geht Gl. (5.37) in Ojas-Regel Gl. (5.32) über. Aber, Gl. (5.37) stellt bezüglich der Nebenbedingung einen allgemeineren Fall im Vergleich zu Gl. (5.32) dar, denn es wird nicht davon ausgegangen, dass die Nebenbedingung erfüllt ist. Vielmehr wird durch den ersten Term auf der rechten Seite von Gl. (5.37) ein Newtonsches Näherungsverfahren[9] zur Berechnung der Nullstelle von der Funktion $f(\hat{\mathbf{v}}_1) = C^2 - \hat{\mathbf{v}}_1^H \hat{\mathbf{v}}_1$ realisiert. Für den reellwertigen, skalaren Fall entspricht dies dem so genannten Babylonischen Wurzelziehen[10], wenn also die Nullstelle von $f(a) = a^2 - \xi$ gesucht wird mit $a \in \mathbb{R}, a > 0$. Die iterative Berechnung der Quadratwurzel von ξ mit dem Iterationsindex κ lautet

$$a_{\kappa+1} = a_\kappa - f(a_\kappa) \left(\frac{\partial f(a_\kappa)}{\partial a_\kappa} \right)^{-1} \tag{5.38}$$

$$= \frac{\xi + a_\kappa^2}{2a_\kappa^2} a_\kappa. \tag{5.39}$$

Das Iterations-Verfahren Gl. (5.39) konvergiert asymptotisch mit quadratischer Konvergenzordnung gegen $\lim_{\kappa \to \infty} a_\kappa = \sqrt{\xi}$. Vergleicht man Gl. (5.39) mit dem linken Term der rechten Seite von Gl. (5.37), so ist zu erkennen, dass die Norm der iterativ berechneten Filterkoeffizienten durch das Newton-Verfahren auf dem Wert C gehalten werden bzw. in einer nahen Umgebung von diesem. Dieses Verhalten führt zu einer erhöhten Stabilität im Vergleich zu Gl. (5.32), was durch Simulationen zum Konvergenzverhalten in [WHU05] gezeigt werden konnte (siehe auch Anhang E.2).

5.1.5 RLS-Ähnliche Konvergenz

Mit der Voraussetzung $\hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1} = C^2 = 1$ soll eine Betrachtung zur Konvergenzbeschleunigung mittels iterationsabhängiger Schrittweite folgen. Diese Betrachtung ist angelehnt an die rekursive Kleinste-Quadrate-Methode (engl. *Recursive Least Squares*, RLS) [Yan95, DK96, CA03]. Üblicherweise wird beim RLS-Algorithmus ein gewünschtes Signal durch ein Eingangssignal mittels Transversalfilterung rekonstruiert. Im Falle der Bestimmung des gesuchten Eigenvektors bedeutet dies jedoch, dass $\hat{\mathbf{v}}_{1,\kappa} \hat{\mathbf{v}}_{1,\kappa}^H \mathbf{X}$ das Eingangssignal $\mathbf{X}$ so gut wie möglich rekonstruiert. Daher lautet die Kostenfunktion anstatt dessen

$$J(\hat{\mathbf{v}}_{1,\kappa}) = E\left\{ ||\mathbf{X} - \hat{\mathbf{v}}_{1,\kappa} \hat{\mathbf{v}}_{1,\kappa}^H \mathbf{X}||^2 \right\} \simeq \sum_{i=1}^{\kappa} \alpha^{\kappa-i} ||\mathbf{X}_i - \hat{\mathbf{v}}_{1,\kappa} \hat{\mathbf{v}}_{1,\kappa}^H \mathbf{X}_i||^2, \tag{5.40}$$

[9] Das Newtonsche Näherungsverfahren wird auch Newton-Raphsonsche Methode genannt.

[10] Das Babylonische Wurzelziehen ist auch bekannt als Heronverfahren nach Heron von Alexandria. Es kann sehr effizient auf digitalen Signalprozessoren eingesetzt werden [AL05].

wobei α einen Glättungsfaktor darstellt mit $0 < \alpha \leq 1$. Benutzt man die Projektionsapproximation $\hat{\mathbf{v}}_{1,\kappa}^H \mathbf{X}_i \approx Y_i$ lässt sich schließlich schreiben

$$J'(\hat{\mathbf{v}}_{1,\kappa}) = \sum_{i=1}^{\kappa} \alpha^{\kappa-i} ||\mathbf{X}_i - \hat{\mathbf{v}}_{1,\kappa} Y_i||^2. \tag{5.41}$$

Der zu Null gesetzte Gradientenvektor von Gl. (5.41) ergibt die optimalen Filterkoeffizienten

$$\hat{\mathbf{v}}_{1,\kappa} = \hat{\mathbf{\Phi}}_{\mathbf{X}Y,\kappa} \hat{\phi}_{YY,\kappa}^{-1}, \tag{5.42}$$

wobei

$$\hat{\mathbf{\Phi}}_{\mathbf{X}Y,\kappa} = \sum_{i=1}^{\kappa} \alpha^{\kappa-i} \mathbf{X}_i Y_i^* = \alpha \hat{\mathbf{\Phi}}_{\mathbf{X}Y,\kappa-1} + \mathbf{X}_\kappa Y_\kappa^*, \tag{5.43}$$

$$\hat{\phi}_{YY,\kappa} = \sum_{i=1}^{\kappa} \alpha^{\kappa-i} Y_i Y_i^* = \alpha \hat{\phi}_{YY,\kappa-1} + Y_\kappa Y_\kappa^*. \tag{5.44}$$

und die Startwerte definiert sind zu

$$\hat{\mathbf{\Phi}}_{\mathbf{X}Y,0} := \mathbf{0} \quad \hat{\phi}_{YY,0} := 0. \tag{5.45}$$

Mit der rekursiven Berechnung von $\hat{\phi}_{YY,\kappa}^{-1}$ und $\hat{\mathbf{\Phi}}_{\mathbf{X}Y,\kappa}$ mittels Matrix Inversion Lemma und der allgemeinen Vorgehensweise zur Bestimmung der RLS-Filterkoeffizienten nach [Hay02] kann schließlich geschrieben werden

$$\hat{\mathbf{v}}_{1,\kappa} = \hat{\mathbf{v}}_{1,\kappa-1} + \hat{\phi}_{YY,\kappa}^{-1} Y_\kappa^* \left(\mathbf{X}_\kappa - Y_\kappa \hat{\mathbf{v}}_{1,\kappa-1} \right). \tag{5.46}$$

Gl. (5.46) zeichnet sich durch den iterationsabhängigen Faktor $\hat{\phi}_{YY,\kappa}^{-1}$ aus, welcher als iterationsabhängige Schrittweite interpretiert werden kann. Bei der Wahl von $0 < \alpha < 1$ verschwindet diese für große κ, da $\hat{\phi}_{YY,\kappa}$ in Gl. (5.44) stetig anwächst. Dies ist zwar für die asymptotische Konvergenz wünschenswert, für die Anwendung zum akustischen *Beamforming* jedoch ungeeignet. Hier ist ja gerade das Verfolgen eines sich ändernden Eigenvektors $\mathbf{v}_1$ wünschenswert. Weitere Untersuchungen bezüglich der Schrittweite sind im Anhang E.2 zu finden.

Als Fazit der Ergebnisse Gl. (5.36), Gl. (5.37), Gl. (5.46) und Gl. (E.11) können zwei Algorithmen für das stochastische Gradientenverfahren angegeben werden; zur Lösung des speziellen Eigenwertproblems mittels Gradientenverfahren und gleichmäßiger Gewichtung (S-Grad-GG):

Algorithmus 3 (S-Grad-GG) *Es gilt $\tilde{\mu}_0^{-1} := 0$. Wähle die Fenstergröße N, eine Glättungskonstante α, den Schrittweitefaktor ρ, den Constraint C und einen Startvektor $\hat{\mathbf{v}}_{1,0} \in \mathbb{C}^M$. Berechne für $\kappa = 1, 2, \ldots$*

$$
\begin{aligned}
Y_\kappa &:= \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{X}_\kappa \\
\tilde{\mu}_\kappa^{-1} &:= \alpha \tilde{\mu}_{\kappa-1}^{-1} + (1-\alpha)|Y_\kappa|^2 \\
\mu_\kappa &:= \tilde{\mu}_\kappa \rho C^2 \\
\hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}^{(\mathrm{GG})} &:= \begin{cases} \dfrac{\kappa-1}{\kappa} \hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa-1}^{(\mathrm{GG})} + \dfrac{1}{\kappa} \mathbf{X}_\kappa \mathbf{X}_\kappa^H & \textit{falls } 1 \le \kappa \le N \\ \hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa-1}^{(\mathrm{GG})} + \dfrac{1}{N} \left(\mathbf{X}_\kappa \mathbf{X}_\kappa^H - \mathbf{X}_{\kappa-N} \mathbf{X}_{\kappa-N}^H \right) & \textit{sonst} \end{cases} \\
Q &:= \hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1} \\
\mathbf{a} &:= \hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}^{(\mathrm{GG})} \hat{\mathbf{v}}_{1,\kappa-1} \\
\hat{\mathbf{v}}_{1,\kappa} &:= \frac{C^2+Q}{2Q} \hat{\mathbf{v}}_{1,\kappa-1} + \mu_\kappa \left(\mathbf{a} - \frac{\hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{a}}{Q} \hat{\mathbf{v}}_{1,\kappa-1} \right)
\end{aligned}
$$

sowie für das spezielle Eigenwertproblem mittels Gradientenverfahren und instantaner Schätzung der Kreuzleistungsdichten (S-Grad-IS):

Algorithmus 4 (S-Grad-IS) *Es gilt $\tilde{\mu}_0^{-1} := 0$. Wähle die Glättungskonstante α, den Schrittweitefaktor ρ, den Constraint C und einen Startvektor $\hat{\mathbf{v}}_{1,0} \in \mathbb{C}^M$. Berechne für $\kappa = 1, 2, \ldots$*

$$
\begin{aligned}
Y_\kappa &:= \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{X}_\kappa \\
\tilde{\mu}_\kappa^{-1} &:= \alpha \tilde{\mu}_{\kappa-1}^{-1} + (1-\alpha)|Y_\kappa|^2 \\
\mu_\kappa &:= \tilde{\mu}_\kappa \rho C^2 \\
Q &:= \hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1} \\
\hat{\mathbf{v}}_{1,\kappa} &:= \frac{C^2+Q}{2Q} \hat{\mathbf{v}}_{1,\kappa-1} + \mu_\kappa Y_\kappa^* \left(\mathbf{X}_\kappa - \frac{Y_\kappa}{Q} \hat{\mathbf{v}}_{1,\kappa-1} \right).
\end{aligned}
$$

Der Parameter C^2 aus der Randbedingung ist wegen der Allgemeinheit eingeführt und der Faktor ρ, mit $0,05 < \rho < 0,5$ soll die Sicherstellung der Konvergenz gewährleisten.

5.1.6 Simulationen zum speziellen Eigenwertproblem

In diesem Abschnitt werden die Konvergenzgeschwindigkeiten des neuen Gradientenverfahrens und der Potenzmethode mit simulierten akustischen Eingangsdaten miteinander verglichen. Das betrachtete Quellsignal hat hier eine zeitliche Länge von ca. 4 Sekunden und wird nach *Szenario-1* für $M = 5$ Mikrophonsignale erzeugt. Mit einer Blocklänge von $L = 256$ und einem Vorschub von $B = 128$ ergibt dies $l_x = 382$ zu verarbeitende Blöcke. Zu beachten ist hierbei, dass die Sprache nach einer sehr kurzen Pause von 0,15 Sekunden, also von 14 Blöcken einsetzt.

Zunächst erfolgt eine Untersuchung von Algorithmus 1 (S-PM-GG) und Algorithmus 3 (S-Grad-GG) hinsichtlich des relativen Fehlers der geschätzten Filterkoeffizienten zu dem wahren Koeffizientenvektor und des erreichten SNRs. Für beide Verfahren gilt $N > l_x$, so dass über die gesamte Länge eine gleichgewichtete Glättung der Kreuzleistungsdichten erfolgt. Weiterhin gilt jeweils für die Initialisierung $\hat{\mathbf{v}}_{1,0} = 1/\sqrt{5} \cdot (1,\ 1,\ 1,\ 1,\ 1)^T$. Bei dem Gradientenverfahren wurde $C = 1$, $\alpha = 0,98$ und $\rho = 0,1$ gesetzt. Nun soll auch wieder die Schreibweise mit der diskreten Frequenzkomponente Ω_k für die Vektoren verwendet werden. Zu jedem Iterationszeitpunkt wird für das aktuell geschätzte $\hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}^{(\mathrm{GG})}(\Omega_k)$ der wahre gesuchte

Eigenvektor $\mathbf{v}_{1,\kappa}(\Omega_k)$ bestimmt. Dieses optimale Verfahren wird mit (S-Opt-GG) gekennzeichnet. Dadurch lässt sich ein relativer Fehler pro Frequenzkomponente definieren zu

$$e(\hat{\mathbf{v}}_{1,\kappa}(\Omega_k)) = \left|\left| \frac{\mathbf{v}_{1,\kappa}(\Omega_k)}{v_{1,1,\kappa}(\Omega_k)} - \frac{\hat{\mathbf{v}}_{1,\kappa}(\Omega_k)}{\hat{v}_{1,1,\kappa}(\Omega_k)} \right|\right| \cdot \left|\left| \frac{\mathbf{v}_{1,\kappa}(\Omega_k)}{v_{1,1,\kappa}(\Omega_k)} \right|\right|^{-1}. \tag{5.47}$$

Damit der Fehler eindeutig ist, wurden die Vektoren in Gl. (5.47) jeweils auf die erste Komponente $\hat{v}_{1,1,\kappa}(\Omega_k)$ bzw. $v_{1,1,\kappa}(\Omega_k)$ normiert. Über alle Frequenzen gemittelt ergibt sich dann der mittlere Fehler

$$\bar{e}(\hat{\mathbf{v}}_{1,\kappa}) = \frac{1}{L} \sum_{k=1}^{L} e(\hat{\mathbf{v}}_{1,\kappa}(\Omega_k)). \tag{5.48}$$

Da das letztendlich wahrgenommene Ergebnis des akustischen *Beamformings* nicht der relative Fehler Gl. (5.48) ist, sondern die Verbesserung des Sprachsignals, soll noch ein frequenzabhängiger asymptotischer SNR-Gewinn nach der Filterung definiert werden

$$\mathrm{SNRG}_{\kappa}(\Omega_k) = \frac{\hat{\mathbf{v}}_{1,\kappa}^H(\Omega_k)\hat{\mathbf{\Phi}}_{\mathbf{SS},l_s}^{(\mathrm{GG})}(\Omega_k)\hat{\mathbf{v}}_{1,\kappa}(\Omega_k)}{\hat{\mathbf{v}}_{1,\kappa}^H(\Omega_k)\hat{\mathbf{\Phi}}_{\mathbf{NN},l_x}^{(\mathrm{GG})}(\Omega_k)\hat{\mathbf{v}}_{1,\kappa}(\Omega_k)} \cdot \frac{\mathrm{Spur}\{\hat{\mathbf{\Phi}}_{\mathbf{NN},l_x}^{(\mathrm{GG})}(\Omega_k)\}}{\mathrm{Spur}\{\hat{\mathbf{\Phi}}_{\mathbf{SS},l_s}^{(\mathrm{GG})}(\Omega_k)\}}. \tag{5.49}$$

In Gl. (5.49) ist mit $\hat{\mathbf{\Phi}}_{\mathbf{SS},l_s}^{(\mathrm{GG})}(\Omega_k)$ die Matrix der Kreuzleistungsdichten des reinen Sprachsignals bezeichnet, die über l_s Blöcke gleichmäßig gewichtet ermittelt wurde. Entsprechend Gl. (5.48) ergibt sich ein asymptotischer SNR-Gewinn gemittelt über alle Frequenzen

$$\overline{\mathrm{SNRG}}_{\kappa} := 10 \cdot \log_{10} \left(\frac{1}{L} \sum_{k=1}^{L} \mathrm{SNRG}_{\kappa}(\Omega_k) \right) \mathrm{dB}. \tag{5.50}$$

In Bild 5.1 sind beispielhafte Verläufe für den Fehler Gl. (5.48) und den asymptotischen SNR-Gewinn Gl. (5.50) für den Fall von lediglich unkorreliertem, weißen Rauschen als Störsignal mit einem SNR pro Eingangssignal von 25 dB dargestellt. In Bild 5.1 (a) und (b) sind diese Verläufe für eine Nachhallzeit von $T_{60} = 0{,}05\,\mathrm{s}$ und in Bild 5.1 (c) und (d) für $T_{60} = 0{,}5\,\mathrm{s}$ zu sehen.

Wird dem mehrkanaligen Sprachsignal noch diffuses Tiefpassrauschen mit einem SNR von 5 dB überlagert, so sind die Ergebnisse in Bild 5.2 zu erreichen. An den repräsentativen Verläufen in den Bildern 5.1 und 5.2 ist klar zu erkennen, dass der gesuchte Eigenvektor gefunden wird, und das schon nach wenigen Iterationsschritten. Da zu Beginn erst einige Signalblöcke zur Schätzung der Kreuzleistungsdichten benötigt werden, ergibt sich ein gewisser Einschwingvorgang, der jeweils besonders an dem Fehler $\bar{e}(\hat{\mathbf{v}}_{1,\kappa})$ zu erkennen ist. Für die Potenzmethode liegen die Kurven für den asymptotischen SNR-Gewinn nahezu auf den optimal ermittelten Verläufen. Bei dem Gradientenverfahren ist eine kleine Verzögerung zu erkennen, die jedoch bei der gewählten Abtastrate und Blocklänge im Bereich von unter 100 ms liegt.

Als letztes sollen noch Verläufe zur Konvergenzgeschwindigkeit präsentiert werden, welche nicht aus einer gleichmäßig gewichteten Schätzung der Matrix $\hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}(\Omega_k)$ hervorgehen, sondern für die Potenzmethode aus einer exponentiellen Glättung nach Algorithmus 2 (S-PM-EG) und für das Gradientenverfahren durch eine instantane Schätzung nach Algorithmus 4 (S-Grad-IS). Das zugrundeliegende Sprachsignal soll aus zwei Sequenzen bestehen. Für die erste ist die Sprecherrichtung wieder 45° wie in den Experimenten für die Bilder 5.1 und 5.2. In der zweiten Sequenz wechselt die Sprecherrichtung nach einer sehr kurzen Pause auf 0°.

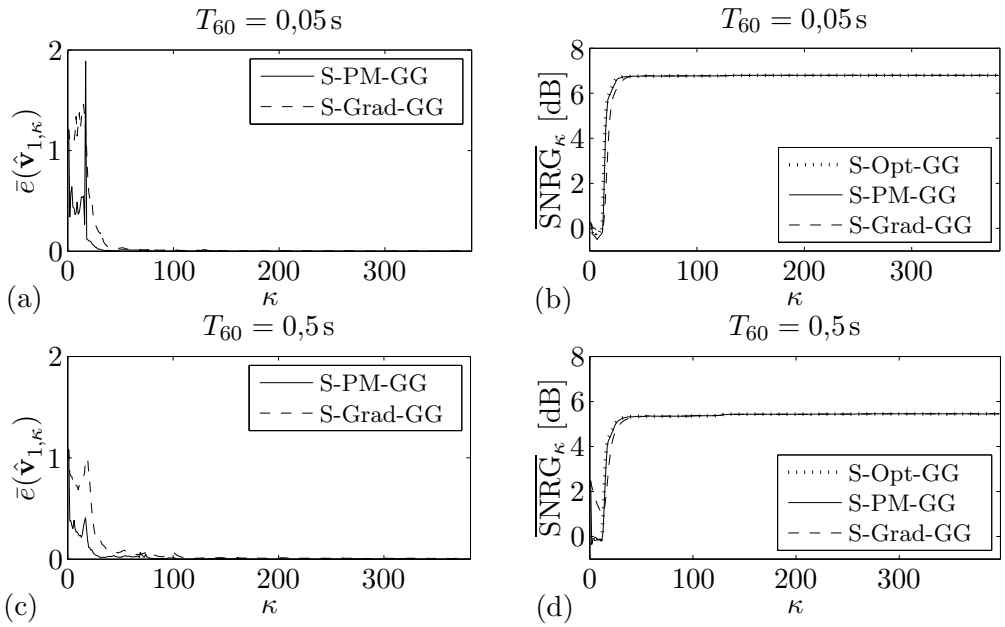

Bild 5.1: Mittlerer Adaptionsfehler und SNR-Gewinn für Algorithmus 1 (S-PM-GG) und Algorithmus 3 (S-Grad-GG) bei unkorreliertem weißen Rauschen als Störsignal.

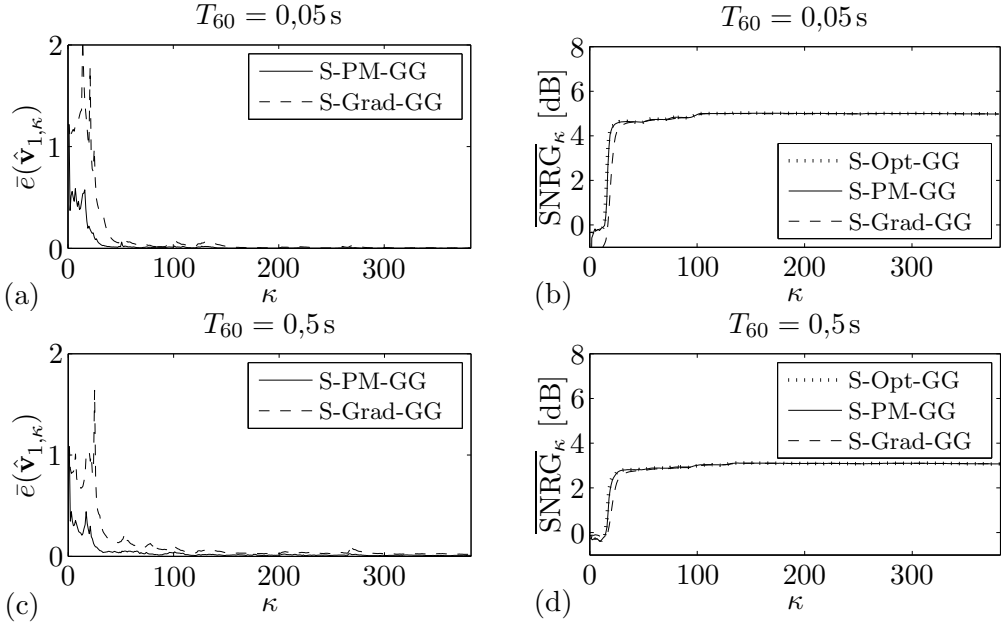

Bild 5.2: Mittlerer Adaptionsfehler und SNR-Gewinn für Algorithmus 1 (S-PM-GG) und Algorithmus 3 (S-Grad-GG) bei diffusem Tiefpassrauschen und additivem unkorrelierten weißen Rauschen als Störsignal.

Die Vektoren wurden wieder jeweils mit $\hat{\mathbf{v}}_{1,0} = 1/\sqrt{5} \cdot (1,\ 1,\ 1,\ 1,\ 1)^T$ initialisiert und die Werte für die weiteren Parameter wurden wie folgt gewählt: $C = 1$, $\alpha = 0,98$ und $\rho = 0,1$.

In Bild 5.3 sind exemplarische Verläufe für den SNR-Gewinn bei rein unkorrelierten additiven Störsignalen mit einem SNR von 25 dB zu sehen; (a) für eine Nachhallzeit von $T_{60} = 0{,}05\,\mathrm{s}$ und (b) für $T_{60} = 0{,}5\,\mathrm{s}$. Zusätzlich sind die SNR-Verläufe dargestellt, welche sich bei der optimalen Bestimmung der Eigenvektoren mit einer gleichmäßig gewichteten Schätzung der Matrizen $\hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},\kappa}(\Omega_k)$ ergeben, die jedoch zu Beginn der zweiten Sprachsequenz neu initia-

lisiert wurden. An den Ergebnissen in Bild 5.3 sind deutlich die Sprünge zu erkennen, die sich durch den Richtungswechsel bei $\kappa = 380$ ergeben. Beide Verfahren, Algorithmus 2 (S-PM-EG) und Algorithmus 4 (S-Grad-IS) folgen recht gut den optimalen Verläufen, wobei für Algorithmus 4 (S-Grad-IS) die Abweichung minimal größer ist.

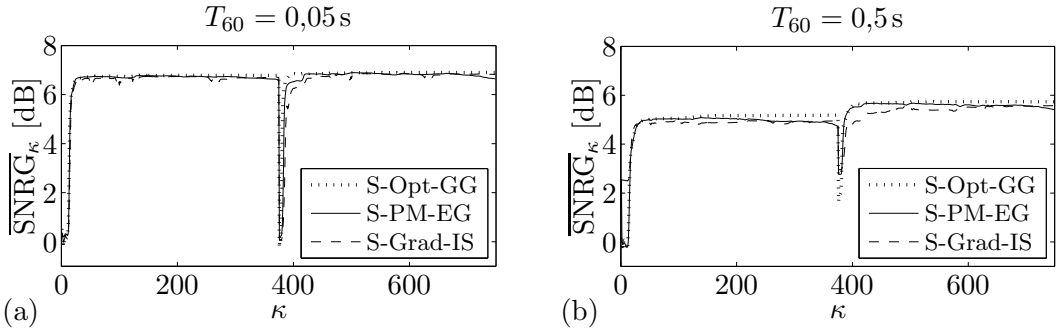

Bild 5.3: SNR-Gewinn für Algorithmus 2 (S-PM-EG) und Algorithmus 4 (S-Grad-IS) bei unkorreliertem weißen Rauschen als Störsignal und einem Wechsel der Sprecherrichtung bei $\kappa = 380$.

Wird den beiden Sprachsequenzen zusätzlich zum unkorrelierten weißen Rauschen noch eine additive diffuse Störung mit einem SNR von 5 dB überlagert, so ergeben sich die beispielhaften Verläufe in Bild 5.4. Aufgrund des recht hohen Störanteils im Eingangssignal sind nun die Schwankungen bezüglich des SNR-Gewinns deutlich ausgeprägter. Dennoch ist gut zu erkennen, dass beide Algorithmen dem optimalen Verlauf folgen und insbesondere auf den abrupten Wechsel der Sprecherrichtung reagiert wird.

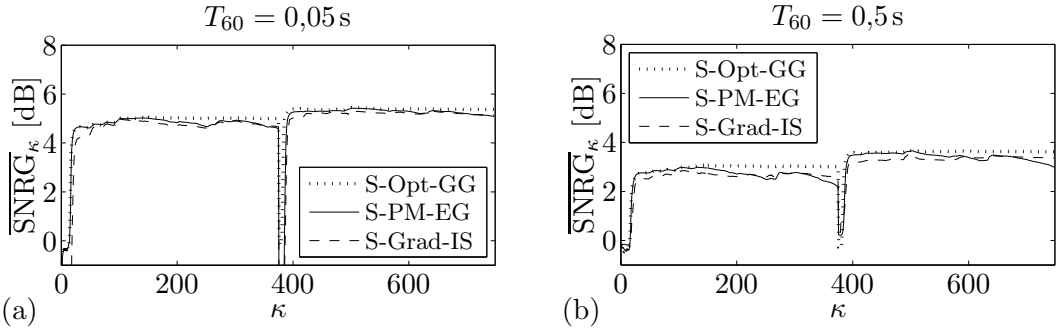

Bild 5.4: SNR-Gewinn für Algorithmus 2 (S-PM-EG) und Algorithmus 4 (S-Grad-IS) bei diffusem Tiefpassrauschen und additivem unkorrelierten weißen Rauschen als Störsignal und einem Wechsel der Sprecherrichtung bei $\kappa = 380$.

Als Fazit lässt sich an dieser Stelle sagen, dass trotz der höheren Schwankungen im SNR-Gewinn das neuartige Gradientenverfahren mit instantaner Schätzung der Kreuzleistungsdichten gemäß Algorithmus 4 (S-Grad-IS) ein sehr schnelles und robustes Verfahren zur Ermittlung und Verfolgung des gesuchten Eigenvektors darstellt. Da hier keinerlei Matrix-Operationen benötigt werden, ist die Komplexität linear in M und somit eine Potenz geringer als die Komplexität der Potenzmethode gemäß Algorithmus 2 (S-PM-EG). Ein weiterer Vorteil ist die einfache Vermeidung von zyklischen Effekten bei der Anwendung des Gradientenverfahrens, welche bisher bei der Auflistung von Algorithmus 4 (S-Grad-IS) außer acht gelassen wurden. In der letztendlichen Implementierung zur mehrkanaligen Sprachsignalverbesserung sind jedoch noch drei Aspekte berücksichtigt worden [Shy92]:

- Die Mikrophonsignale werden mittels *Overlap-Save*-Verfahrens mit den Filterkoeffizien-

ten gefiltert.

- Die Subtraktion in dem Fehlerterm $\mathbf{X}_\kappa(\Omega_k) - Y_\kappa(\Omega_k)/(\hat{\mathbf{v}}_{1,\kappa-1}^H(\Omega_k)\hat{\mathbf{v}}_{1,\kappa-1}(\Omega_k)) \cdot \hat{\mathbf{v}}_{1,\kappa-1}(\Omega_k)$ wird im Zeitbereich durchgeführt.
- Der Gradiententerm, also die gesamte Änderung der Filterkoeffizienten von einem Iterationsschritt zum nächsten, wird im Zeitbereich für die zweite Hälfte der Impulsantworten auf Null gesetzt.

5.2 Allgemeines Eigenwertproblem

In diesem Abschnitt wird die Kreuzleistungsdichtematrix der Störung $\mathbf{\Phi_{NN}}$ beim Eigenwertproblem mit berücksichtigt

$$\mathbf{\Phi_{XX}}\mathbf{v}_i = \lambda_i \mathbf{\Phi_{NN}}\mathbf{v}_i, \tag{5.51}$$

mit den hermiteschen, positiv definiten Matrizen $\mathbf{\Phi_{XX}}, \mathbf{\Phi_{NN}} \in \mathbb{C}^{M\times M}$. Es soll wieder von normierten Eigenvektoren mit $||\mathbf{v}_i|| = 1\ \forall i$ ausgegangen werden. Die Eigenwerte sind wiederum reellwertig und positiv, weshalb auch hier folgende Sortierung gelten soll:

$$\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_M \geq 0. \tag{5.52}$$

Gesucht wird ein Eigenvektor $\mathbf{v}_1$ korrespondierend zum größten Eigenwert λ_1. Dafür sollen im Folgenden zum einen Gradientenverfahren verwendet werden, die direkt die Matrizen $\mathbf{\Phi_{XX}}$ und $\mathbf{\Phi_{NN}}$ benötigen. Zum anderen kommen Fixpunktverfahren zum Einsatz, welche die Berechnung der Inversen von $\mathbf{\Phi_{NN}}$ voraussetzen und somit das allgemeine in ein spezielles Eigenwertproblem umformen.

5.2.1 Potenzmethode und Projektionsapproximation

Das allgemeine Eigenwertproblem Gl. (5.51) kann in folgendes spezielles Eigenwertproblem umgeschrieben werden

$$\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{\Phi_{XX}}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \tag{5.53}$$

so dass äquivalent zu Gl. (5.6) für den gesuchten Eigenvektor gilt

$$\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{\Phi_{XX}}\mathbf{v}_1 = \lambda_1 \mathbf{v}_1 \tag{5.54}$$

$$\frac{\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{\Phi_{XX}}\mathbf{v}_1}{||\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{\Phi_{XX}}\mathbf{v}_1||} = \mathbf{v}_1. \tag{5.55}$$

Die iterative Lösung ergibt sich entsprechend zu

$$\hat{\mathbf{v}}_{1,\kappa} = \frac{\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{\Phi_{XX}}\hat{\mathbf{v}}_{1,\kappa-1}}{||\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{\Phi_{XX}}\hat{\mathbf{v}}_{1,\kappa-1}||} \quad \kappa = 1, 2, 3, \ldots \tag{5.56}$$

mit dem Startvektor $\hat{\mathbf{v}}_{1,0} = \sum_{i=1}^{M} c_i \mathbf{v}_i,\ c_i \in \mathbb{C}, c_1 \neq 0$. Für die Konvergenz gilt entsprechend den Überlegungen in Abschnitt 5.1.1, dass die Konvergenzrate wieder maßgeblich durch das Verhältnis λ_2/λ_1 bestimmt wird und die Folge $\{\hat{\mathbf{v}}_{1,\kappa}\}_{\kappa\in\mathbb{N}}$ in Gl. (5.56) linear gegen $c_1/|c_1|\mathbf{v}_1$ konvergiert. Zusätzlich zu den Startwerten c_i hängt der Iterationsfehler von einem Iterationsschritt zum nächsten noch von den Eigenwerten von $\mathbf{\Phi_{NN}}$ ab. Je kleiner das Verhältnis zwischen dem größten und dem kleinsten Eigenwert der Matrix $\mathbf{\Phi_{NN}}$ ist, je ähnlicher $\mathbf{\Phi_{NN}}$

also der Einheitsmatrix wird, desto genauer wird im Allgemeinen die Näherung $\hat{\mathbf{v}}_{1,\kappa}$ für den Schritt κ [Krü07].

Beim Einsatz für das akustische *Beamforming* sind nun zunächst die Matrizen $\mathbf{\Phi}_{\mathbf{NN}}^{-1}$ und $\mathbf{\Phi}_{\mathbf{XX}}$ zu schätzen. Für die Inverse der Störleistungsdichten soll eine rekursive Glättung[11] nach Gl. (A.29) verwendet werden. Diese Schätzung wird zu Zeitpunkten durchgeführt, in denen nur das Störsignal an den Sensoren vorliegt. *Vice versa* wird $\mathbf{\Phi}_{\mathbf{XX}}$ geschätzt, während Sprachaktivität vorliegt. Während dieser Sequenzen erfolgt ebenfalls wechselseitig die Iteration des gesuchten Eigenvektors. Die Schätzung $\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}$ ist während dieser Zeiten unverändert und soll daher keinen Iterationsindex tragen. Es sollen nun zwei Algorithmen für die stochastische Potenzmethode angegeben werden; zur Lösung des allgemeinen Eigenwertproblems mittels Potenzmethode und gleichmäßiger Gewichtung (A-PM-GG):

Algorithmus 5 (A-PM-GG) *Gegeben sei* $\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}$. *Setze* $\boldsymbol{A}_0^{(\mathrm{GG})} := \mathbf{0}$. *Wähle die Fenstergröße* N *und einen Startvektor* $\hat{\mathbf{v}}_{1,0} \in \mathbb{C}^M$. *Berechne für* $\kappa = 1, 2, \ldots$

$$
\begin{aligned}
\mathbf{b}_\kappa &:= \hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}\mathbf{X}_\kappa \\
\boldsymbol{A}_\kappa^{(\mathrm{GG})} &:= \begin{cases} \dfrac{\kappa-1}{\kappa}\boldsymbol{A}_{\kappa-1}^{(\mathrm{GG})} + \dfrac{1}{\kappa}\mathbf{b}_\kappa\mathbf{X}_\kappa^H & \textit{falls } 1 \le \kappa \le N \\ \boldsymbol{A}_{\kappa-1}^{(\mathrm{GG})} + \dfrac{1}{N}\left(\mathbf{b}_\kappa\mathbf{X}_\kappa^H - \mathbf{b}_{\kappa-N}\mathbf{X}_{\kappa-N}^H\right) & \textit{sonst} \end{cases} \\
\mathbf{a} &:= \boldsymbol{A}_\kappa^{(\mathrm{GG})}\hat{\mathbf{v}}_{1,\kappa-1} \\
\hat{\mathbf{v}}_{1,\kappa} &:= \frac{\mathbf{a}}{||\mathbf{a}||}
\end{aligned}
$$

sowie für das allgemeine Eigenwertproblem mittels Potenzmethode und exponentieller Gewichtung (A-PM-EG):

Algorithmus 6 (A-PM-EG) *Gegeben sei* $\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}$. *Setze* $\boldsymbol{A}_0^{(\mathrm{EG})} := \mathbf{0}$. *Wähle eine Glättungskonstante* α *und einen Startvektor* $\hat{\mathbf{v}}_{1,0} \in \mathbb{C}^M$. *Berechne für* $\kappa = 1, 2, \ldots$

$$
\begin{aligned}
\mathbf{b}_\kappa &:= \hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}\mathbf{X}_\kappa \\
\boldsymbol{A}_\kappa^{(\mathrm{EG})} &:= \alpha\boldsymbol{A}_{\kappa-1}^{(\mathrm{EG})} + (1-\alpha)\mathbf{b}_\kappa\mathbf{X}_\kappa^H \\
\mathbf{a} &:= \boldsymbol{A}_\kappa^{(\mathrm{EG})}\hat{\mathbf{v}}_{1,\kappa-1} \\
\hat{\mathbf{v}}_{1,\kappa} &:= \frac{\mathbf{a}}{||\mathbf{a}||}
\end{aligned}
$$

Der Rechenaufwand der Potenzmethode lässt sich wiederum nach der Methode der Projektionsapproximation gemäß des Vorgehens in 5.1.2 reduzieren. Dafür sollen zwei Algorithmen angegeben werden; zur Lösung des allgemeinen Eigenwertproblems mittels Projektionsapproximation und gleichmäßiger Gewichtung (A-PA-GG)

[11] Alternativ zur iterativen Berechnung der Inversen von $\mathbf{\Phi}_{\mathbf{NN}}$ kann das allgemeine Eigenwertproblem auch durch eine Cholesky-Zerlegung von $\mathbf{\Phi}_{\mathbf{NN}}$ in ein spezielles Eigenwertproblem umgeformt werden.

Algorithmus 7 (A-PA-GG) *Gegeben sei* $\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}$. *Setze* $\boldsymbol{p}_0^{(\mathrm{GG})} := \mathbf{0}$. *Wähle die Fenstergröße* N *und einen Startvektor* $\hat{\mathbf{v}}_{1,0} \in \mathbb{C}^M$. *Berechne für* $\kappa = 1, 2, \ldots$

$$
\begin{aligned}
Y_\kappa^* &:= \mathbf{X}_\kappa^H \hat{\mathbf{v}}_{1,\kappa-1} \\
\boldsymbol{p}_\kappa^{(\mathrm{GG})} &:= \begin{cases} \dfrac{\kappa-1}{\kappa}\boldsymbol{p}_{\kappa-1}^{(\mathrm{GG})} + \dfrac{1}{\kappa}\mathbf{X}_\kappa Y_\kappa^* & \text{falls } 1 \le \kappa \le N \\ \boldsymbol{p}_{\kappa-1}^{(\mathrm{GG})} + \dfrac{1}{N}\left(\mathbf{X}_\kappa Y_\kappa^* - \mathbf{X}_{\kappa-N} Y_{\kappa-N}^*\right) & \text{sonst} \end{cases} \\
\mathbf{a} &:= \hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1} \mathbf{p}_\kappa^{(\mathrm{GG})} \\
\hat{\mathbf{v}}_{1,\kappa} &:= \frac{\mathbf{a}}{||\mathbf{a}||}
\end{aligned}
$$

sowie für das allgemeinen Eigenwertproblem mittels Projektionsapproximation und exponentieller Gewichtung (A-PA-EG):

Algorithmus 8 (A-PA-EG) *Gegeben sei* $\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}$. *Setze* $\boldsymbol{p}_0^{(\mathrm{EG})} := \mathbf{0}$. *Wähle eine Glättungskonstante* α *und einen Startvektor* $\hat{\mathbf{v}}_{1,0} \in \mathbb{C}^M$. *Berechne für* $\kappa = 1, 2, \ldots$

$$
\begin{aligned}
Y_\kappa^* &:= \mathbf{X}_\kappa^H \hat{\mathbf{v}}_{1,\kappa-1} \\
\boldsymbol{p}_\kappa^{(\mathrm{EG})} &:= \alpha \boldsymbol{p}_{\kappa-1}^{(\mathrm{EG})} + (1-\alpha)\mathbf{X}_\kappa Y_\kappa^* \\
\mathbf{a} &:= \hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1} \mathbf{p}_\kappa^{(\mathrm{EG})} \\
\hat{\mathbf{v}}_{1,\kappa} &:= \frac{\mathbf{a}}{||\mathbf{a}||}
\end{aligned}
$$

Für die Konvergenz von Algorithmus 7 (A-PA-GG) gelten prinzipiell die gleichen Überlegungen wie für Algorithmus 5 (A-PM-GG).

Weitere untersuchte Verfahren für die iterative Berechnung des Eigenvektors zum größten Eigenwert wie z. B. das Minimierungsverfahren mittels einer Quasi-Newton-Methode [MRP96] oder das RLS-basierte Verfahren [YXYZ06] weisen ein äquivalentes oder schlechteres Adaptionsverhalten als die hier gezeigten Algorithmen auf [Krü07].

5.2.2 Neuartiges Gradientenverfahren

Das im Weiteren vorgestellte Gradientenverfahren basiert auf den gleichen Herleitungsschritten wie jenes in Abschnitt 5.1.4, jedoch mit dem Unterschied, dass die Kreuzleistungsdichtematrix der Störung mit einbezogen wird. Für das Maximierungsproblem bedeutet dies

$$\max_{\mathbf{v}^H} \mathbf{v}^H \mathbf{\Phi}_{\mathbf{XX}} \mathbf{v} \quad \text{unter der Randbed.} \quad \mathbf{v}^H \mathbf{\Phi}_{\mathbf{NN}} \mathbf{v} = C^2, \quad C \in \mathbb{R}^+. \tag{5.57}$$

Mit dem Lagrange-Multiplikator β kann dann eine Kostenfunktion angegeben werden

$$J(\mathbf{v}, \beta) = \mathbf{v}^H \mathbf{\Phi}_{\mathbf{XX}} \mathbf{v} + \beta(\mathbf{v}^H \mathbf{\Phi}_{\mathbf{NN}} \mathbf{v} - C^2), \tag{5.58}$$

deren Gradientenvektor

$$\nabla_{\mathbf{v}} J(\mathbf{v}, \beta) = 2\mathbf{\Phi}_{\mathbf{XX}} \mathbf{v} + 2\beta \mathbf{\Phi}_{\mathbf{NN}} \mathbf{v}, \tag{5.59}$$

in die Iterationsgleichung für $\hat{\mathbf{v}}_{1,\kappa}$ mittels deterministischem Gradientenanstieg

$$\hat{\mathbf{v}}_{1,\kappa} = \hat{\mathbf{v}}_{1,\kappa-1} + \frac{\mu}{2} \nabla_{\mathbf{v}} J(\mathbf{v}, \beta)\Big|_{\mathbf{v}=\hat{\mathbf{v}}_{1,\kappa-1}} \tag{5.60}$$

einzusetzen ist. Um den Lagrange-Multiplikator zu berechnen wird nun gefordert, dass die Nebenbedingung für den Iterationsschritt κ eingehalten bleibt

$$\hat{\mathbf{v}}_{1,\kappa}^H \mathbf{\Phi}_{\mathbf{NN}} \hat{\mathbf{v}}_{1,\kappa} \stackrel{!}{=} C^2. \tag{5.61}$$

Nach dem Einsetzen von Gl. (5.60) in Gl. (5.61) und der Verwendung von Gl. (5.59) ergibt sich unter Vernachlässigung der Terme quadratisch in μ die Näherung

$$\begin{aligned} C^2 \approx\ & \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1} + \mu \hat{\mathbf{v}}_{1,\kappa-1}^H \left(\mathbf{\Phi_{XX}} \mathbf{\Phi_{NN}} + \mathbf{\Phi_{NN}} \mathbf{\Phi_{XX}} \right) \hat{\mathbf{v}}_{1,\kappa-1} \\ & + 2\beta\mu \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{NN}} \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1}. \end{aligned} \tag{5.62}$$

Zur kürzeren Schreibweise soll die Definition

$$\mathbf{\Phi}^{(\mathbf{XN})} = \mathbf{\Phi_{XX}} \mathbf{\Phi_{NN}} + \mathbf{\Phi_{NN}} \mathbf{\Phi_{XX}} \tag{5.63}$$

eingeführt werden, welche in die nach β aufgelöste Näherung Gl. (5.62) eingesetzt wird

$$\beta \approx \frac{C^2 - \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1} - \mu \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi}^{(\mathbf{XN})} \hat{\mathbf{v}}_{1,\kappa-1}}{2\mu \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{NN}} \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1}}. \tag{5.64}$$

Unter Ausnutzung von Gl. (5.64) kann Gl. (5.60) mit Gl. (5.59) angegeben werden zu

$$\begin{aligned} \hat{\mathbf{v}}_{1,\kappa} = & \left[\mathbf{I} + \frac{C^2 - \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1}}{2\hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{NN}} \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1}} \mathbf{\Phi_{NN}} \right] \hat{\mathbf{v}}_{1,\kappa-1} \\ & + \mu \left(\mathbf{\Phi_{XX}} \hat{\mathbf{v}}_{1,\kappa-1} - \frac{\hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi}^{(\mathbf{XN})} \hat{\mathbf{v}}_{1,\kappa-1}}{2\hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{NN}} \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1}} \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1} \right). \end{aligned} \tag{5.65}$$

Definiert man weiter

$$\mathbf{D}_{\kappa-1} = \mathbf{I} + \frac{C^2 - \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1}}{2\hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{NN}} \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1}} \mathbf{\Phi_{NN}} \qquad \xi_{\kappa-1} = \frac{\hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi}^{(\mathbf{XN})} \hat{\mathbf{v}}_{1,\kappa-1}}{2\hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{NN}} \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1}}, \tag{5.66}$$

so ergibt sich für Gl. (5.65)

$$\hat{\mathbf{v}}_{1,\kappa} = \mathbf{D}_{\kappa-1} \hat{\mathbf{v}}_{1,\kappa-1} + \mu \left(\mathbf{\Phi_{XX}} \hat{\mathbf{v}}_{1,\kappa-1} - \xi_{\kappa-1} \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1} \right). \tag{5.67}$$

Die Interpretation von Gl. (5.67) ist nun zweierlei. Zum einen sorgt die Matrix $\mathbf{D}_{\kappa-1}$ für die Einhaltung der Randbedingung und wird gerade zur Einheitsmatrix wenn diese erfüllt ist. Zum anderen bewirkt die Zielfunktion $\xi_{\kappa-1}$ eine Art Steuerung der Anteile der beiden Vektoren in den Klammern auf der rechten Seite von Gl. (5.67). Denn durch das positive Vorzeichen von $\mathbf{\Phi_{XX}} \hat{\mathbf{v}}_{1,\kappa-1}$ strebt der Vektor in die Richtung, die $\hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{XX}} \hat{\mathbf{v}}_{1,\kappa-1}$ maximiert. Und das negative Vorzeichen vor dem Ausdruck $\mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1}$ bewirkt eine Verstärkung des Vektors der Richtung, welche $\hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1}$ minimiert. Beide Ausdrücke sind bekanntlich im Gleichgewicht, wenn die Zielfunktion dem größten Eigenwert λ_1 und $\hat{\mathbf{v}}_{1,\kappa-1}$ dem korrespondierenden Eigenvektor $\mathbf{v}_1$ entspricht.

Bei zahlreichen Experimenten hat sich herausgestellt, dass die Matrix $\mathbf{D}_{\kappa-1}$ zwar für eine sehr gute Einhaltung der Randbedingung sorgt, allerdings auch zu schwankenden Abweichungen von dem optimalen Vektor führen kann. Dieses Verhalten wird durch die drehende Wirkung von $\mathbf{D}_{\kappa-1}$ verursacht, also hin zu der dominanten Komponente von $\mathbf{\Phi_{NN}}$ bei Unterschreitung der Randbedingung und entsprechend weg von der dominanten Richtung von $\mathbf{\Phi_{NN}}$ bei Überschreitung der Randbedingung. Daher wird eine heuristische Änderung von Gl. (5.67) vorgenommen und $\mathbf{D}_{\kappa-1}$ ersetzt durch

$$\widetilde{D}_{\kappa-1} := \frac{C^2 + \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1}}{2\hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1}}, \tag{5.68}$$

also durch lediglich einen Skalar der eine reine Längenänderung von $\hat{\mathbf{v}}_{1,\kappa-1}$ bewirkt. Die Verwendung von Gl. (5.68) ist motiviert durch die Erkenntnisse aus Abschnitt 5.1.4.

Als nächstes soll die Bedeutung der Zielfunktion $\xi_{\kappa-1}$ erläutert werden. Dazu wird das verallgemeinerte Eigenwertproblem aus Gl. (5.51) auf beiden Seiten mit $\mathbf{\Phi_{NN}}$ von links multipliziert und umgestellt

$$\lambda_i = \frac{\mathbf{v}_i^H \mathbf{\Phi_{NN}} \mathbf{\Phi_{XX}} \mathbf{v}_i}{\mathbf{v}_i^H \mathbf{\Phi_{NN}} \mathbf{\Phi_{NN}} \mathbf{v}_i}. \tag{5.69}$$

Für beliebige Vektoren $\mathbf{v}$ in Gl. (5.69) ergibt sich an Stelle von λ_i ein komplexwertiger Skalar, dessen Realteil die Form

$$\Re\left\{\frac{\mathbf{v}^H \mathbf{\Phi_{NN}} \mathbf{\Phi_{XX}} \mathbf{v}}{\mathbf{v}^H \mathbf{\Phi_{NN}} \mathbf{\Phi_{NN}} \mathbf{v}}\right\} = \frac{\mathbf{v}^H \mathbf{\Phi}^{(\mathbf{XN})} \mathbf{v}}{2\mathbf{v}^H \mathbf{\Phi_{NN}} \mathbf{\Phi_{NN}} \mathbf{v}} = \xi(\mathbf{v}) \tag{5.70}$$

annimmt und man erkennt beim Vergleich mit Gl. (5.66), dass Gl. (5.70) einen zur Zielfunktion $\xi_{\kappa-1}$ äquivalenten Ausdruck darstellt. Mit dem Rayleigh Quotienten

$$r(\mathbf{v}) = \frac{\mathbf{v}^H \mathbf{\Phi_{XX}} \mathbf{v}}{\mathbf{v}^H \mathbf{\Phi_{NN}} \mathbf{v}} \tag{5.71}$$

kann zwar für $\mathbf{v} = \mathbf{v}_i$ gefolgert werden, dass $\xi(\mathbf{v}_i) = r(\mathbf{v}_i)$ gilt, für beliebige $\mathbf{v}$ ist jedoch der theoretische Zusammenhang sehr schwierig zu zeigen. Daher sollen anhand von Monte-Carlo-Simulationen Streudiagramme (engl. *Scatterplot*) zur graphischen Darstellung der Wertepaare $\xi(\mathbf{v})$ und $r(\mathbf{v})$ präsentiert werden. Grundlage hierfür ist wieder das *Szenario-2* und die Matrizen $\mathbf{\Phi_{XX}}$ und $\mathbf{\Phi_{NN}}$ sollen optimal geschätzt sein. Dann kann für zufällig gewählte Vektoren $\mathbf{v}$ der sich ergebende Wert $\xi(\mathbf{v})$ über $r(\mathbf{v})$ als Punkt in ein kartesisches Koordinatensystem eingetragen werden. In Bild 5.5 sind für unterschiedliche Frequenzen Streudiagramme abgebildet. Die Nachhallzeit liegt bei $T_{60} = 0{,}05\,\mathrm{s}$ und die Anzahl der zufällig gezogenen komplexen Vektoren betrug 1000. Das Bild 5.6 zeigt die Streudiagramme für unterschiedliche Frequenzen bei

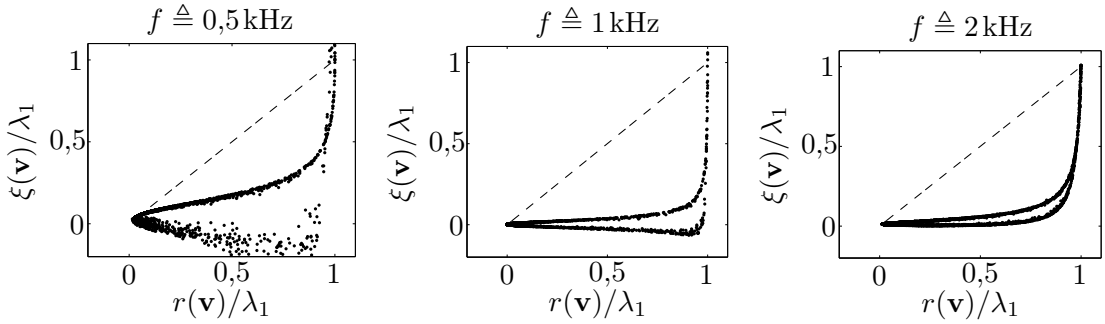

Bild 5.5: Streudiagramme für unterschiedliche Frequenzen ausgewertet für das *Szenario-2* mit optimal bestimmten Matrizen und einer Nachhallzeit von $T_{60} = 0{,}05\,\mathrm{s}$.

einer Nachhallzeit von $T_{60} = 0{,}5\,\mathrm{s}$ und die Anzahl der zufällig gezogenen komplexen Vektoren betrug wieder 1000. Interessant an den beispielhaften Ergebnissen in den Diagrammen Bild 5.5 und Bild 5.6 ist, dass keine eindeutige Aussage über den Zusammenhang von $\xi(\mathbf{v})$ und $r(\mathbf{v})$ gemacht werden kann. Es lassen sich lediglich zwei Tendenzen ausmachen. Zum einen fällt die Abweichung zwischen $\xi(\mathbf{v})$ und $r(\mathbf{v})$ bei steigender Nachhallzeit meistens kleiner aus, und zum anderen nähert sich $\xi(\mathbf{v})$ dem Wert von $r(\mathbf{v})$ in der Regel von unten an, wenn sich der ausgewertete Vektor der dominanten Komponente $\mathbf{v}_1$ nähert. Die Interpretationen dieser Tendenzen ist, dass in Gl. (5.67) die Maximierung durch $\mathbf{\Phi_{XX}} \hat{\mathbf{v}}_{1,\kappa-1}$ gegenüber der Minimierung mittels $\xi_{\kappa-1} \mathbf{\Phi_{NN}} \hat{\mathbf{v}}_{1,\kappa-1}$ dominiert. Und zwar um so stärker, je "schärfer" der Sprecher

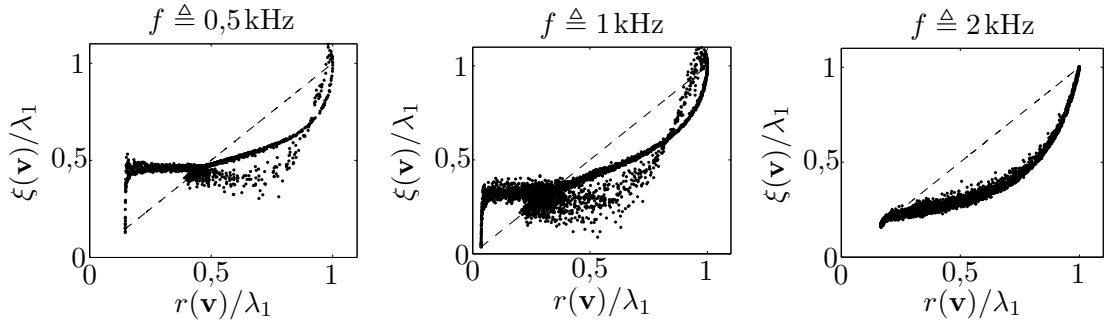

Bild 5.6: Streudiagramme für unterschiedliche Frequenzen ausgewertet für das *Szenario-2* mit optimal bestimmten Matrizen und einer Nachhallzeit von $T_{60} = 0{,}5\,\mathrm{s}$.

von der Störquelle zu trennen ist (also für geringe Nachhallzeiten). Bei der Iteration von $\hat{\mathbf{v}}_{1,\kappa}$ wird also prinzipiell die Ausgangsleistung gegeben die Statistik des Mischsignals von Sprache und Störung schneller maximiert als die Leistung des gefilterten Störsignals minimiert wird.

Bezüglich der Wahl der Schrittweite sind im Anhang E.2 Abschätzungen aufgrund von Simulationen zu finden. Als Ergebnis dieser Experimente soll ein Wertebereich für eine Schrittweite angegeben werden:

$$\mu_\kappa = \frac{\rho}{r_\kappa}, \qquad 0,05 < \rho < 1. \tag{5.72}$$

In Gl. (5.72) ist mit ρ zwar ein frei wählbarer doch während der Adaption konstanter Schrittweitefaktor bezeichnet. Der Parameter r_κ stellt den Rayleigh Quotienten zum aktuellen Iterationsschritt dar. Weiterhin wird die KLDS-Matrix der Störung in einer normierten Version verwendet: $\widetilde{\hat{\mathbf{\Phi}}}_{\mathbf{NN}} = \hat{\mathbf{\Phi}}_{\mathbf{NN}}/\hat{\sigma}_N^2$, mit $\hat{\sigma}_N^2 = \mathrm{Spur}\{\hat{\mathbf{\Phi}}_{\mathbf{NN}}\}/\mathrm{M}$.

Abschließend soll ein Algorithmus zur Lösung des allgemeinen Eigenwertproblems mittels Gradientenverfahren und gleichmäßiger Gewichtung in zwei Varianten angegeben werden; mit der Zielfunktion wie sie sich nach der Herleitung in Gl. (5.67) (A-Grad-GG) ergibt und alternativ mit dem aktuellen Rayleigh Quotienten als Zielfunktion (A-RQgrad-GG):

Algorithmus 9 (A-Grad-GG) und (A-RQgrad-GG) *Gegeben sei* $\hat{\mathbf{\Phi}}_{\mathbf{NN}}$ *und somit* $\widetilde{\hat{\mathbf{\Phi}}}_{\mathbf{NN}} = \hat{\mathbf{\Phi}}_{\mathbf{NN}}/\hat{\sigma}_N^2$. *Setze* $\hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},0} := \mathbf{0}$. *Wähle die Fenstergröße* N, *den Schrittweitefaktor* ρ, *den Constraint* C *und einen Startvektor* $\hat{\mathbf{v}}_{1,0} \in \mathbb{C}^M$. *Berechne für* $\kappa = 1, 2, \ldots$

$$\begin{aligned}
\hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},\kappa} &:= \begin{cases} \frac{\kappa-1}{\kappa}\hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},\kappa-1} + \frac{1}{\kappa}\mathbf{X}_\kappa\mathbf{X}_\kappa^H & \textit{falls } 1 \le \kappa \le N \\ \hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},\kappa-1} + \frac{1}{N}\left(\mathbf{X}_\kappa\mathbf{X}_\kappa^H - \mathbf{X}_{\kappa-N}\mathbf{X}_{\kappa-N}^H\right) & \textit{sonst} \end{cases} \\
\mathbf{a} &:= \hat{\mathbf{\Phi}}^{(\mathrm{GG})}_{\mathbf{XX},\kappa}\hat{\mathbf{v}}_{1,\kappa-1} \\
\mathbf{b} &:= \widetilde{\hat{\mathbf{\Phi}}}_{\mathbf{NN}}\hat{\mathbf{v}}_{1,\kappa-1} \\
Q &:= \hat{\mathbf{v}}_{1,\kappa-1}^H\mathbf{b} \\
r &:= \frac{\hat{\mathbf{v}}_{1,\kappa-1}^H\mathbf{a}}{Q} \\
\xi &:= \begin{cases} r & \textit{für Algorithmus (A-RQgrad-GG)} \\ \Re\{\frac{\mathbf{a}^H\mathbf{b}}{\mathbf{b}^H\mathbf{b}}\} & \textit{für Algorithmus (A-Grad-GG)} \end{cases} \\
\hat{\mathbf{v}}_{1,\kappa} &:= \frac{C^2+Q}{2Q}\hat{\mathbf{v}}_{1,\kappa-1} + \frac{\rho}{r}(\mathbf{a} - \xi\mathbf{b})
\end{aligned}$$

Für das allgemeine Eigenwertproblem mittels Gradientenverfahren und instantaner Schätzung der Kreuzleistungsdichten ergibt sich (A-Grad-IS) und alternativ mit dem aktuellen

Rayleigh Quotienten als Zielfunktion (A-RQgrad-IS):

Algorithmus 10 (A-Grad-IS) und (A-RQgrad-IS) *Gegeben sei* $\hat{\mathbf{\Phi}}_{\mathbf{NN}}$ *und somit* $\widetilde{\mathbf{\Phi}}_{\mathbf{NN}} = \hat{\mathbf{\Phi}}_{\mathbf{NN}}/\hat{\sigma}_N^2$. *Setze* $P_0 := \mathbf{0}$. *Wähle die Glättungskonstante* α, *den Schrittweitefaktor* ρ, *den Constraint* C *und einen Startvektor* $\hat{\mathbf{v}}_{1,0} \in \mathbb{C}^M$. *Berechne für* $\kappa = 1, 2, \ldots$

$$
\begin{aligned}
Y_\kappa &:= \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{X}_\kappa \\
P_\kappa &:= \alpha P_{\kappa-1} + (1-\alpha)|Y_\kappa|^2 \\
\mathbf{b} &:= \widetilde{\mathbf{\Phi}}_{\mathbf{NN}} \hat{\mathbf{v}}_{1,\kappa-1} \\
Q &:= \hat{\mathbf{v}}_{1,\kappa-1}^H \mathbf{b} \\
r &:= \frac{P_\kappa}{Q} \\
\xi &:= \begin{cases} r & \text{für Algorithmus (A-RQgrad-IS)} \\ \Re\{\frac{Y_\kappa \mathbf{X}_\kappa^H \mathbf{b}}{\mathbf{b}^H \mathbf{b}}\} & \text{für Algorithmus (A-Grad-IS)} \end{cases} \\
\hat{\mathbf{v}}_{1,\kappa} &:= \frac{C^2+Q}{2Q} \hat{\mathbf{v}}_{1,\kappa-1} + \frac{\rho}{r}(Y_\kappa^* \mathbf{X}_\kappa - \xi \mathbf{b})
\end{aligned}
$$

5.2.3 Simulationen zum allgemeinen Eigenwertproblem

Beispiele zum Konvergenzverhalten der im letzten Abschnitt vorgestellten Verfahren sollen im Folgenden präsentiert werden. Das betrachtete Sprachsignal hat eine zeitliche Länge von ca. 4 Sekunden, mit dessen Hilfe $M = 5$ Mikrophonsignale nach *Szenario-2* für unterschiedliche Nachhallzeiten erzeugt werden. Das Sprachsignal fällt also aus einer Richtung von 45° und das gerichtete Tiefpassrauschen unter einem Winkel von -20° auf die Sensoren ein, wobei das Tiefpassrauschen mit einem SNR von 5 dB hinzugemischt wurde. Zusätzlich sind den einzelnen Signalpfaden jeweils unkorreliertes weißes Rauschen mit einem SNR pro Eingangssignal von 25 dB überlagert. Die Blocklänge beträgt wieder $L = 256$, der Vorschub $B = 128$ und die Anzahl zu verarbeitenden Blöcke ergibt $l_x = 382$.

Die untersuchten Verfahren sind zunächst Algorithmus 5 (A-PM-GG), Algorithmus 7 (A-PA-GG) und die beiden Varianten Algorithmus 9 (A-Grad-GG)/(A-RQgrad-GG). Es gilt $N > l_x$, so dass über die gesamte Länge eine gleichgewichtete Glättung der Kreuzleistungsdichten erfolgt und die Initialisierung ist zu $\hat{\mathbf{v}}_{1,0} = 1/\sqrt{5} \cdot (1,\ 1,\ 1,\ 1,\ 1)^T$ gewählt. Für die Gradientenverfahren wird $\widetilde{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k) = M\hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k)/\text{Spur}\{\hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k)\}$ eingesetzt, wodurch mit $C = 1/32$ eine Reduzierung der Störleistung vom Eingang zum Ausgang um ca. 15 dB festgelegt wird. Weiterhin wurde der Schrittweitefaktor zu $\rho = 0,6$ gesetzt. In Bild 5.7 ist der Fehler Gl. (5.48) und der asymptotische SNR-Gewinn Gl. (5.50) aufgetragen: in (a) und (b) für eine Nachhallzeit von $T_{60} = 0{,}05\,\text{s}$ und in (c) und (d) für $T_{60} = 0{,}5\,\text{s}$. Aus Übersichtlichkeitsgründen wird auf den Verlauf des optimalen Ergebnisses verzichtet.

An den Ergebnissen in Bild 5.7 sind drei Eigenschaften festzustellen:

- Der Unterschied zwischen dem Verfahren mit Projektionsapproximation und der Potenzmethode ist sehr gering. Die Approximation Gl. (5.17) ist also zulässig und führt kaum zu Einbußen.
- Bei den zwei Varianten des Gradientenverfahrens ist kein wesentlicher Unterschied zu erkennen.
- Trotz eines Fehlers $\bar{e}(\hat{\mathbf{v}}_{1,\kappa}) \neq 0$ kann der SNR-Gewinn nahezu konvergiert sein.

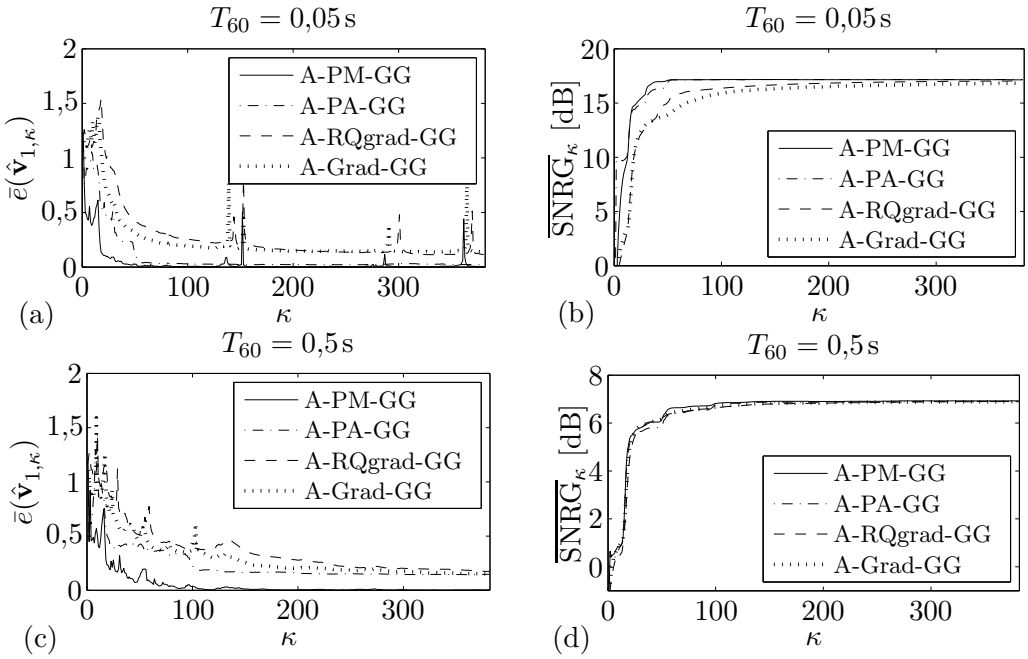

Bild 5.7: Mittlerer Adaptionsfehler und SNR-Gewinn für Algorithmus 5 (A-PM-GG), Algorithmus 7 (A-PA-GG) und die beiden Varianten Algorithmus 9 (A-Grad-GG)/(A-RQgrad-GG) bei gerichtetem Tiefpassrauschen und additivem unkorrelierten weißen Rauschen als Störsignal.

- Für die Gradientenverfahren stellt sich eine schnellere Konvergenz mit steigender Nachhallzeit ein.

Um einer sich ändernden Statistik zu folgen – hervorgerufen etwa durch einen sich bewegenden Sprecher – ist wieder beim Einsatz für das akustische *Beamforming* von der gleichgewichteten Glättung abzusehen. Es wird daher die exponentielle Glättung für die Potenzmethode Algorithmus 6 (A-PM-EG) und für die Projektionsapproximation aus Algorithmus 8 (A-PA-EG) verwendet. Für die beiden Gradientenverfahren kommt die instantane Schätzung in Algorithmus 10 (A-Grad-IS)/(A-RQgrad-IS) zum Einsatz. Das zugrundeliegende Sprachsignal soll aus zwei Sequenzen bestehen. Für die erste ist die Sprecherrichtung wieder 45° wie in den Experimenten in Bild 5.7, und in der zweiten Sequenz wechselt die Sprecherrichtung nach einer sehr kurzen Pause auf 0°. Die Initialisierung der Vektoren wurde zu $\hat{\mathbf{v}}_{1,0} = 1/\sqrt{5} \cdot (1,\ 1,\ 1,\ 1,\ 1)^T$ gewählt und die Werte der weiteren Parameter betrugen $C = 1/32$, $\alpha = 0,98$ und $\rho = 0,6$. Die KLDS-Matrix der Störung kam in der normierten Form $\widetilde{\hat{\mathbf{\Phi}}}_{\mathbf{NN}}(\Omega_k) = M\hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k)/\mathrm{Spur}\{\hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k)\}$ zum Einsatz. Exemplarische Ergebnisse dieser Anordnung sind in Bild 5.8 dargestellt; links für eine Nachhallzeit von $T_{60} = 0{,}05\,\mathrm{s}$ und rechts für $T_{60} = 0{,}5\,\mathrm{s}$. Es zeigt sich hierbei ein deutlicher Unterschied zwischen den Gradientenverfahren und den Fixpunktalgorithmen. Obschon die Schrittweite für die Gradientenverfahren relativ hoch gewählt wurde, ist die Konvergenzgeschwindigkeit im Vergleich zur Potenzmethode und dem Verfahren mit Projektionsapproximation signifikant geringer, insbesondere bei niedrigen Nachhallzeiten. Wird zusätzlich zur gerichteten Störung noch diffuses Rauschen hinzuaddiert, so fällt der Unterschied im Konvergenzverhalten umso geringer aus, je höher der Anteil des diffusen Rauschens im Verhältnis zur gerichteten Störung ist.

Abschließend lässt sich bezüglich der vorgestellten Verfahren zur adaptiven Berechnung des Eigenvektors korrespondierend zum größten Eigenwert eines allgemeinen Eigenwertproblems folgern, dass zwar die Komplexität $\mathcal{O}(M^2)$ für die Potenzmethode und die Gradien-

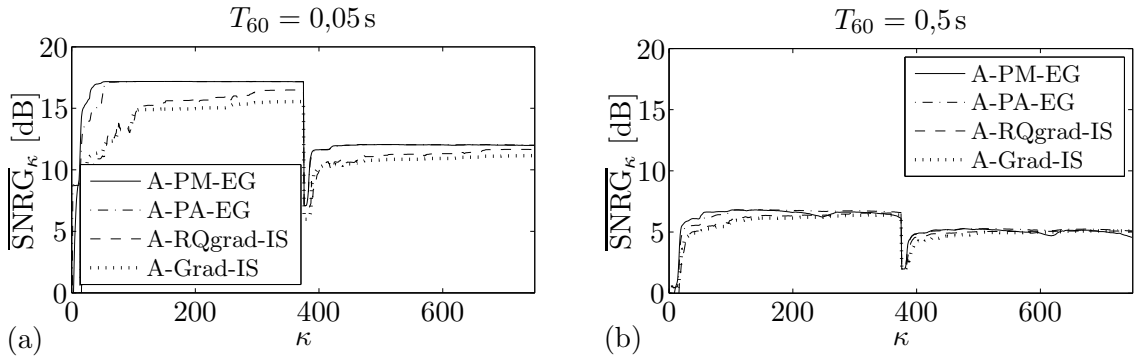

Bild 5.8: SNR-Gewinn für Algorithmus 6 (A-PM-EG), Algorithmus 8 (A-PA-EG) und die beiden Varianten Algorithmus 10 (A-Grad-IS)/(A-RQgrad-IS) bei einem Sprecherwechsel und stationärer Störung bestehend aus gerichtetem Tiefpassrauschen und additivem unkorrelierten, weißen Rauschen .

tenverfahren gleich sind, das Adaptionsverhalten der Potenzmethode jedoch deutlich besser ausfällt. Einen geringeren Rechenaufwand erfordert der Algorithmus mittels Projektionsapproximation bei sehr ähnlichem Verhalten wie die Potenzmethode. Bei den beiden Varianten des Gradientenverfahrens lässt sich keine eindeutige Präferenz aussprechen. Der größte Nachteil bei beiden Varianten liegt darin, dass ein geeigneter Schrittweitefaktor gewählt werden muss. Dieser Nachteil sollte dann in Kauf genommen werden, wenn $\hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k)$ ebenfalls durch eine instantane Schätzung approximiert wird und die Rechenkomplexität dadurch um eine Potenz geringer ausfällt, also linear in M ist. In [Mor04] werden z. B. zweistufige Gradientenverfahren für den Anwendungsbereich in der Mobilfunktechnik beschrieben die auf einer instantanen Schätzung der KLDS-Matrix der Störung basieren.

Zum Einsatz der Fixpunktverfahren für das *Beamforming* muss bei der letztendlichen Implementierung im Frequenzbereich und der Nutzung des *Overlap-Save*-Verfahrens auf die Vermeidung der zyklischen Faltung geachtet werden. Ein mögliches Vorgehen ist hierbei:

- Pro Verarbeitungskanal ist die Blocklänge bzw. FFT-Länge L mit dem Vorschub B, so dass L zu filternde Spektralkomponenten anfallen, für die also L Filterkoeffizienten berechnet werden.
- Nach der Rücktransformation der Filterkoeffizienten in den Zeitbereich werden $L - B$ Filterkoeffizienten pro Signalpfad herausgeschnitten, mit Nullen auf die Länge L aufgefüllt und wieder in den Frequenzbereich transformiert.

Anmerkungen zu $\hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k)$ Die KLDS-Matrix kann mittels einer exponentiellen Glättung in den Sprachpausen geschätzt werden. Da angenommen wird, dass sich die Statistik der Störung nur langsam ändert, kann diese Schätzung auch während Sprachaktivität als gültig erachtet werden. Durch die exponentielle Glättung wird gewährleistet, dass langsame Änderungen von Sprachpause zu Sprachpause erfassbar sind.

Bei unkorrelierten Störanteilen und bei diffusem Rauschen sind die jeweiligen Strukturen von $\mathbf{\Phi}_{\mathbf{NN}}(\Omega_k)$ gegeben durch die Einheitsmatrix bzw. die mit si-Termen gefüllte Kohärenzmatrix. Aber diese beiden Strukturen ergeben sich erst nach der Erwartungswert-Bildung über eine große Menge von Eingangsdaten. Hingegen gilt für den Anteil einer gerichteten Störung $\mathbf{N}_m(\Omega_k)\mathbf{N}_m^H(\Omega_k) = |N_{c,m}(\Omega_k)|^2\mathbf{A}(\Omega_k)\mathbf{A}^H(\Omega_k)$, mit $\mathbf{A}(\Omega_k)$ als Übertragungsfunktion der gerichteten Störung. Jeder Block m enthält somit bereits die Information über die Struktur von $\mathbf{\Phi}_{\mathbf{NN}}$, welcher als Beitrag in das exponentiell geglättete $\hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k)$ eingeht. Dies

bedeutet also, dass nur ein paar Blöcke für eine gute Schätzung notwendig sind. Bei steigender Nachhallzeit kommen bei rein gerichteten Störungen noch diffuse Komponenten zu $\hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k)$ aufgrund der dann zu kurzen Blocklänge hinzu. Es sind dann also mehr Eingangsblöcke für eine gute Schätzung notwendig.

Aus Robustheitsgründen sollte generell noch ein Regularisierungsterm der Größenordnung $-30\,\mathrm{dB}$ bis $-40\,\mathrm{dB}$ überlagert werden [Bit02]: $\hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k) := \hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k) + \delta\sigma_N^2(\Omega_k)\mathbf{I}$, mit $0,001 < \delta < 0,0001$.

Anmerkungen zu $\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}(\Omega_k)$ Für Mikrophongruppen mit wenigen Sensoren ist auch eine direkte Inversion von $\hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k)$ am Ende einer Sprachpause denkbar. In dieser Arbeit findet jedoch die rekursive Schätzung nach Gl. (A.29) Anwendung. Dabei kann insbesondere eine geringe Quantisierungsauflösung zu numerischen Problemen führen. Auch hier ist mittels eines zusätzlichen Regularisierungsterms eine Steigerung der Robustheit der Schätzung $\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}(\Omega_k)$ zu erzielen. Allerdings muss hierfür bei der inversen Schätzung zu den Eingangsdaten ein Rauschen hinzuaddiert werden. Eine effiziente Implementierung ist dabei im Frequenzbereich möglich [Fis07]. In jedem Iterationsschritt wird ein M-dimensionaler, komplexer Vektor aus einer Normalverteilung gezogen und entsprechend gewichtet zur ersten Frequenzkomponente der Eingangsdaten hinzuaddiert. Diese Zufallswerte werden dann nach jedem Iterationsschritt in Richtung steigender Frequenzkomponenten verschoben und zu diesen neu gewichtet hinzuaddiert.

Bezüglich der Geschwindigkeit für eine vertrauenswürdige Schätzung gelten die gleichen Überlegungen wie bei der Ermittlung von $\hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k)$.

Anmerkungen zur Nichtstationarität der Sprache Es gelten die oben gemachten Anmerkungen zu $\hat{\mathbf{\Phi}}_{\mathbf{NN}}$ und $\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}$, wobei an dieser Stelle wieder auf die frequenzabhängige Notation verzichtet wird. Diese Matrizen sind während der Adaption von $\hat{\mathbf{v}}_{1,\kappa}$ unverändert, jedoch ist die Varianz der Sprache $\phi_{S_cS_c,\kappa}$ nun abhängig von dem Iterationsschritt, welcher gleichbedeutend mit dem Blockindex ist. Das allgemeine Eigenwertproblem kann somit formuliert werden zu

$$\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}\hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}\hat{\mathbf{v}}_{1,\kappa} = \hat{\lambda}_{1,\kappa}\hat{\mathbf{v}}_{1,\kappa} \tag{5.73}$$

$$\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}\left[\phi_{S_cS_c,\kappa}\mathbf{H}\mathbf{H}^H + \hat{\mathbf{\Phi}}_{\mathbf{NN}}\right]\hat{\mathbf{v}}_{1,\kappa} = \hat{\lambda}_{1,\kappa}\hat{\mathbf{v}}_{1,\kappa}, \tag{5.74}$$

mit der aktuellen Schätzung $\hat{\lambda}_{1,\kappa}$ für den größten Eigenwert. Weiter umgestellt folgt aus Gl. (5.74) schließlich

$$\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}\mathbf{H}\mathbf{H}^H\hat{\mathbf{v}}_{1,\kappa} = \frac{\hat{\lambda}_{1,\kappa} - 1}{\phi_{S_cS_c,\kappa}}\hat{\mathbf{v}}_{1,\kappa}, \quad \text{mit} \quad \phi_{S_cS_c,\kappa} \neq 0. \tag{5.75}$$

An Gl. (5.75) ist zu erkennen, dass die Nichtstationarität der Sprache lediglich die "Länge" des geschätzten Eigenvektors ändert aber nicht dessen "Richtung". Da aber nach jedem Iterationsschritt die Schätzung $\hat{\mathbf{v}}_{1,\kappa}$ auf die Einheitslänge normiert wird, spielt diese Tatsache für das *Beamforming* keine Rolle, solange sich die Position des Sprechers – und damit die Übertragungsfunktion $\mathbf{H}$ – nicht ändert.

5.3 Zusammenfassung

In diesem Kapitel wurden iterative Verfahren zur Bestimmung des Eigenvektors korrespondierend zum größten Eigenwert eines speziellen und des allgemeinen Eigenwertproblems präsentiert und miteinander verglichen. Einerseits waren dies Fixpunktverfahren wie die Potenzmethode und der Algorithmus mittels Projektionsapproximation und andererseits eigenentwickelte Gradientenverfahren.

Die experimentellen Ergebnisse für das allgemeine Eigenwertproblem bezüglich der Konvergenz zeigen eine Überlegenheit der Fixpunktverfahren im Vergleich zu den Gradientenverfahren, insbesondere, da sie unabhängig von Schrittweitefaktoren sind. Daher sollte die Potenzmethode zum Einsatz für das akustische *Beamforming* unter Berücksichtigung der Kreuzleistungsdichtematrix der Störung präferiert werden. Um eine Nachführung der Filterkoeffizienten bei einem sich bewegenden Sprecher zu ermöglichen, ist das stochastische Verfahren Algorithmus 6 (A-PM-EG) mit exponentieller Glättung der KLDS-Matrix der Eingangsdaten einzusetzen. Für die neuartige GSC-Struktur mittels adaptiver Eigenwertzerlegung in Kapitel 8 sollte jedoch das Verfahren Algorithmus 5 (A-PM-GG) verwendet werden, da dort von keinerlei (oder sehr geringen) Sprecherbewegungen während der Adaption ausgegangen wird.

Beim Einsatz eines *Beamformers* mit den optimalen Filterkoeffizienten nach dem Max-SNR-Kriterium in einer "gemäßigten" Umgebung, wenn also außer dem Sprecher keine weiteren dominanten Schallquellen zu erwarten sind, sollte lediglich das spezielle Eigenwertproblem der Kreuzleistungsdichtematrix der Mikrophonsignale gelöst werden. Hier zeigt das neuartige Gradientenverfahren vergleichbare Konvergenzeigenschaften wie die Potenzmethode auf, hat jedoch eine deutlich geringere Rechenkomplexität. Daher kann unter diesen Randbedingungen das eigenentwickelte stochastische Gradientenverfahren Algorithmus 4 (S-Grad-IS) eingesetzt werden. Dieses ist als Erweiterung der bekannten Adaptionsregel nach Oja anzusehen, jedoch im Vergleich zu dieser weist das neue Verfahren eine signifikante Steigerung der Robustheit bezüglich der Stabilität auf, was in den vergleichenden Analysen im Anhang gezeigt werden konnte.

Kapitel 6

Einkanaliges Nachfilter für das Eigenvektor-Beamforming

In Kapitel 4 wurde gezeigt, dass unterschiedliche Optimierungskriterien zu statistisch optimalen Filterkoeffizienten führen, welche sich nur in einem skalaren Faktor unterscheiden. Hierbei zeigt das Max-SNR-Kriterium insbesondere den Vorteil, dass keinerlei Wissen über die geometrische Anordnung zur Bestimmung der Filterkoeffizienten notwendig ist. Diese Koeffizienten können über adaptive Algorithmen zur Lösung eines Eigenwertproblems im Frequenzbereich, wie sie in Kapitel 5 vorgestellt wurden, berechnet werden. Es ergibt sich also der iterativ bestimmte Vektor

$$\hat{\mathbf{v}}_1(\Omega) = \tilde{\mathbf{F}}_{\mathrm{SNR}}(\Omega) = \zeta(\Omega)\mathbf{F}_{\mathrm{SNR}}(\Omega), \qquad \zeta(\Omega) \in \mathbb{C}. \tag{6.1}$$

Die Filterung der mehrkanaligen Eingangsdaten mit einem Eigenvektor korrespondierend zum verallgemeinerten Eigenwertproblem wird als *Generalized Eigenvector* (GEV) *Beamforming* bezeichnet. Bei der Filterung der Eingangsdaten mit einem Eigenvektor korrespondierend zum speziellen Eigenwertproblem hingegen wird hier von *Principal Component Analysis* (PCA) *Beamforming* gesprochen.

Da die Maximierung des frequenzabhängigen Schmalband-SNRs im Allgemeinen zu Verzerrungen des breitbandigen Sprachsignals führt, sollen in diesem Abschnitt Verfahren vorgestellt werden, welche ebendiese Verzerrungen deutlich reduzieren können. Dabei liegt die Grundidee darin, die Filterkoeffizienten mit $w(\Omega)$ so zu normalisieren, dass sie denen des GMVDR *Beamformers* näherungsweise entsprechen:

$$w(\Omega)\hat{\mathbf{v}}_1(\Omega) \approx \mathbf{F}_{\mathrm{GMVDR}}(\Omega), \qquad w(\Omega) \in \mathbb{R}. \tag{6.2}$$

Da also diese Normalisierung für jeden Verarbeitungszweig durchgeführt wird, kann auch synonym von einer einkanaligen Nachfilterung gesprochen werden.

Für den GMVDR *Beamformer* ist das explizite Wissen der Raumübertragungsfunktion notwendig. Die im folgenden beschriebenen Normalisierungsverfahren nutzen jedoch das implizit in den Filterkoeffizienten $\mathbf{F}_{\mathrm{SNR}}(\Omega) = \mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega)$ steckende Wissen über die Raumübertragungsfunktion aus.

Vorgestellt werden sollen drei mögliche Methoden zur Realisierung des Nachfilters[1] $w(\Omega)$ [WHU06a, WHU07]: eine analytische Näherung für den Fall perfekt ermittelter Eigenvekto-

[1] Bei den in dieser Arbeit vorgestellten Verfahren soll versucht werden, die enthallende Wirkung des *Be-*

ren, und zwei weitere Verfahren, die auf Eigenschaften der Richtcharakteristik des *Beamformers* beruhen. Letztere zeichnen sich dadurch aus, dass nicht zwangsläufig von konvergierten Filterkoeffizienten ausgegangen wird.

6.1 Analytische Normalisierung

Um ein unverzerrtes Sprachsignal am Ausgang des *Beamformers* zu erhalten, muss für die Gesamtübertragungsfunktion bestehend aus dem Koeffizientenvektor $\hat{\mathbf{v}}_1$ und aus der Raumübertragungsfunktion[2] $\mathbf{H}(\Omega)$ von der Quelle zu den Sensoren gelten

$$|w^*_{\text{opt}}(\Omega)\hat{\mathbf{v}}_1^H(\Omega)\mathbf{H}(\Omega)| = 1 \tag{6.3}$$

mit dem optimalen Nachfilter[3]

$$|w_{\text{opt}}(\Omega)| = \left|\frac{1}{\hat{\mathbf{v}}_1^H(\Omega)\mathbf{H}(\Omega)}\right| \tag{6.4}$$

$$= \frac{|\zeta(\Omega)|}{\hat{\mathbf{v}}_1^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\hat{\mathbf{v}}_1(\Omega)}. \tag{6.5}$$

Offensichtlich ist weder der Vektor $\mathbf{H}(\Omega)$ in Gl. (6.4) noch der Skalar $\zeta(\Omega)$ in Gl. (6.5) bekannt. Daher wird nun zunächst der Ausdruck $\mathbf{\Phi_{NN}}(\Omega)\hat{\mathbf{v}}_1(\Omega)$ betrachtet

$$||\mathbf{\Phi_{NN}}(\Omega)\hat{\mathbf{v}}_1(\Omega)||^2 = |\zeta(\Omega)|^2||\mathbf{H}(\Omega)||^2 \tag{6.6}$$

und folgende Näherung hinzugenommen

$$||\mathbf{H}(\Omega)||^2 \approx ||\mathbf{d}(\Omega,\theta_t)||^2 = M, \tag{6.7}$$

mit dem Steering Vektor $\mathbf{d}(\Omega,\theta_t)$ aus Gl. (3.34) für ein linear und äquidistant angeordnetes *Array*. Die Näherung Gl. (6.7) ist motiviert durch die Tatsache, dass bei kurzen Filterlängen des GEV *Beamformers* sich bezüglich des Nutzsignals im Wesentlichen ein Ausgleich der Laufzeitdifferenzen der direkten Ausbreitungpfade ausbildet. Das Nachfilter, welches sich analytisch aus Gl. (6.5) und Gl. (6.6), sowie mit Hilfe der Näherung Gl. (6.7) angeben lässt, soll als blinde analytische Normalisierung (BAN) bezeichnet werden:

$$w_{\text{BAN}}(\Omega) = \frac{||\hat{\mathbf{v}}_1^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)||}{\hat{\mathbf{v}}_1^H(\Omega)\mathbf{\Phi_{NN}}(\Omega)\hat{\mathbf{v}}_1(\Omega)\cdot\sqrt{M}}. \tag{6.8}$$

Der Begriff "analytisch" soll darauf hinweisen, dass hier eine geschlossene Lösung bzw. Näherung angegeben werden kann, im Gegensatz zu den noch folgenden Verfahren. "Blind" ist die Normalisierung Gl. (6.8) aufgrund der Tatsache, dass keine Informationen über die Position von den Mikrophonen bzw. den akustischen Quellen enthalten ist. Aufgrund der analytisch zu

amformings beizuhalten. Im Gegensatz dazu wurde in [HUKW08] eine Methode vorgeschlagen, bei der eine Normalisierung auf das Sprachsignal eines Signalpfades hin erfolgt. Die Halleigenschaften dises Pfades sind dann am Ausgang des *Beamformers* wiederzufinden.

[2] Es soll nochmal darauf hingewiesen werden, dass im Abschnitt 3.1 die vereinfachte Schreibweise $\mathbf{H}(\Omega) := \mathbf{H}(\Omega, \mathbf{p}_s, \mathbf{p}_1, ..., \mathbf{p}_M)$ für die mehrkanalige Raumübertragungsfunktion eingeführt wurde, in der die Abhängigkeit von der Position der Schallquelle $\mathbf{p}_s$ und der Mikrophone $\mathbf{p}_i$ im Raum aus Übersichtlichkeitsgründen vernachlässigt ist. Außerdem ergeben sich folglich adaptiv berechnete Filterkoeffizienten, die ebenfalls von den geometrischen Daten abhängen.

[3] An die Phase der herzuleitenden Nachfilter soll keinerlei Bedingung gestellt werden.

berechnenden Normalisierungsfaktoren $w_{\mathrm{BAN}}(\Omega)$ stellt dieser Nachfilterungsalgorithmus zwar ein relativ einfaches Verfahren dar. Der wesentliche Nachteil liegt jedoch in der Tatsache, dass die Koeffizienten $\hat{\mathbf{v}}_1(\Omega)$ exakt bestimmt worden sein müssen, damit Gl. (6.6) zutrifft. Bei dem realen Einsatz des GEV *Beamformers* ist diese Bedingung jedoch aufgrund zeitveränderlicher Verhältnisse nicht immer gewährleistet.

6.2 Statistische Normalisierung

Es soll nun wieder von den optimalen Faktoren aus Gl. (6.4) ausgegangen werden, allerdings jedoch für eine Freifeld-Anordnung

$$|w_{\mathrm{opt}}(\Omega)|\Big|_{\substack{\theta_t=\theta_s\\ \mathbf{H}(\Omega)=\mathbf{d}(\Omega,\theta_s)}} = \left|\frac{1}{\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\theta_s)}\right|. \tag{6.9}$$

Da ein blindes *Beamforming* realisiert werden soll, ist die Richtung θ_s als unbekannt anzunehmen. Daher wird hier ein statistisch motivierter Ansatz zur Schätzung der Sprecherrichtung bzw. der Normalisierungskoeffizienten vorgeschlagen:

$$w_{\mathrm{BSN}}(\Omega) = \frac{1}{\int\limits_{-\pi/2}^{\pi/2} p(\theta;\Omega)|\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\theta)|d\theta}. \tag{6.10}$$

Für die blinde statistische Normalisierung (BSN) Gl. (6.10) ist eine frequenzabhängige Wahrscheinlichkeitsdichtefunktion $p(\theta;\Omega)$ bezüglich der gesuchten Sprecherrichtung eingeführt. Optimaler Weise sollte die Wahrscheinlichkeitsdichtefunktion gleich der entsprechend verschobenen Delta-Distribution[4] sein $p(\theta;\Omega) = \delta(\theta-\theta_s)$, wodurch dann Gl. (6.10) in Gl. (6.9) übergeht.

Da keine weiteren Verfahren zur Bestimmung der Sprecherrichtung verwendet werden sollen, wird das implizite Wissen über die gesuchte Richtung in den Filterkoeffizienten benutzt. Denn für das *Beampattern* sollten folgende Bedingungen gelten

$$\theta_s \approx \underset{\theta}{\operatorname{argmax}}|\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\theta)| \tag{6.11}$$

$$|\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\theta_s)| \gg |\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\theta_{n,i})|, \qquad \forall i \tag{6.12}$$

wobei $\theta_{n,i}$ die Richtung der i-ten Störquelle beschreibt. So kann die räumliche Übertragungsfunktion selbst in normalisierter Form als Wahrscheinlichkeitsdichtefunktion dienen

$$p(\theta;\Omega) = \frac{|\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\theta)|}{\int\limits_{-\pi/2}^{\pi/2} |\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\theta)|d\theta}. \tag{6.13}$$

Mit Gl. (6.13) eingesetzt in Gl. (6.10) ergibt sich schließlich für die blinde statistische

[4] Die Delta-Distribution ist definiert durch $\delta(x) = \begin{cases} 0 & \text{für } x \neq 0 \\ \infty & \text{falls } x = 0 \end{cases}$.

Normalisierung

$$w_{\mathrm{BSN}}(\Omega) = \frac{\int\limits_{-\pi/2}^{\pi/2} |\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\theta)| d\theta}{\int\limits_{-\pi/2}^{\pi/2} |\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\theta)|^2 d\theta}. \tag{6.14}$$

Die Bezeichnung "blind" trifft für das BSN-Verfahren zwar nicht mehr auf die Anordnung der Mikrophongruppe zu, da ja der Mikrophonabstand für den Steering Vektor bekannt sein muss. Aber die Position des Sprechers im Raum ist weiterhin nicht notwendigerweise explizit zu bestimmen.

Es sei noch angemerkt, dass für die Realisierung der blinden statistischen Normalisierung die Integrale in Gl. (6.14) in Summen zu überführen sind und das *Beampattern* für $2N+1$ diskrete Stützstellen auszuwerten ist:

$$w_{\mathrm{BSN}}(\Omega) = \frac{\sum\limits_{i=-N}^{N} |\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\theta_i)|}{\sum\limits_{i=-N}^{N} |\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\theta_i)|^2} \qquad \text{mit } \theta_i = \frac{\pi}{2N} i. \tag{6.15}$$

6.3 Maximum-Normalisierung

Das Nachfilter Gl. (6.14) führt dazu, dass die resultierende räumliche Übertragungsfunktion in Richtung des Sprechers im Mittel für alle Frequenzen gleich ist. Es wird also nicht auf einzelne Werte des *Beampatterns* für bestimmte Richtungen vertraut, sondern auf die Gesamtheit der Übertragungsfunktion. Der Nachteil liegt also in der Mitberücksichtigung von breiten Hauptkeulen für tiefe Frequenzen und *Grating Lobes* für hohe Frequenzen. Zahlreiche experimentelle Untersuchungen und die sehr guten Adaptionseigenschaften der Algorithmen aus dem Abschnitt 5 zeigen jedoch, dass insbesondere die Annahme Gl. (6.11) recht gut eingehalten wird (siehe auch Kapitel 7). Daher soll die instantane frequenzabhängige Richtungsschätzung[5]

$$\hat{\theta}_s(\Omega) = \underset{\theta}{\mathrm{argmax}} |\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\theta)| \tag{6.16}$$

in Gl. (6.9) eingesetzt und diese als Maximum-Normalisierung (MN) bezeichnet werden

$$w_{\mathrm{MN}}(\Omega) = \frac{1}{|\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\hat{\theta}_s(\Omega)|}. \tag{6.17}$$

Die Maximum-Normalisierung ist für $\theta_t = \hat{\theta}_s$ vergleichbar mit der impliziten Normalisierung der MVDR-Filterkoeffizienten Gl. (4.33). Da jedoch die GEV-*Beamformer*-Koeffizienten das Ausgangs-SNR maximieren, ist zu erwarten, dass die Schätzung $\hat{\theta}_s(\Omega)$ fehlerbehaftet ist. Aber für eine explizite Sprecherrichtungsbestimmung mittels eines gesonderten Verfahrens und dem Einsatz des MVDR *Beamformers* ist ebenfalls davon auszugehen, dass θ_t in Gl. (4.33) nur eine gute Schätzung der gesuchten Richtung darstellt.

Für die Maximum-Normalisierung wird nicht mehr die Bezeichnung "blind" verwendet, da eine frequenzabhängige Richtungsbestimmung in Gl. (6.16) vorgenommen wird.

[5] Die Richtungsschätzung mittels Gl. (6.16) wird durch die Maximum-Suche über diskrete Stützstellen $|\hat{\mathbf{v}}_1^H(\Omega)\mathbf{d}(\Omega,\theta_i)|$ realisiert.

6.4 Simulationen zu Normalisierungsverfahren

In diesem Abschnitt soll die Auswirkung der Normalisierungsverfahren für das akustische *Beamforming* veranschaulicht werden. Dazu erfolgt eine Aufteilung der Problemstellung ohne und mit Berücksichtigung der Kreuzleistungsdichten des Störschallfeldes; also in der Implementierung als PCA *Beamformer* für den ersten Fall und entsprechend als GEV *Beamformer* für den zweiten Fall.

6.4.1 PCA Beamforming

Für die experimentellen Ergebnisse zur verallgemeinerten MVDR-Lösung in Abschnitt 4.5 sowie der Herleitung von Verfahren zur Lösung des speziellen Eigenwertproblems in Abschnitt 5.1 wurde nicht auf die Normierung der Filterkoeffizienten eingegangen. Bei der Betrachtung des letztendlichen Ausgangssignals ist diese jedoch sehr wichtig und wird hier für den PCA *Beamformer* mittels BAN-Methode vorgeschlagen.

Da für das PCA *Beamforming* das spezielle Eigenwertproblem gelöst wird, ist die KLDS-Matrix $\mathbf{\Phi_{NN}}(\Omega)$ nicht berücksichtigt bzw. kann gleich der Einheitsmatrix gesetzt werden. Das Nachfilter wird somit zu

$$w_{\mathrm{BAN}}(\Omega) = \frac{1}{\sqrt{M\hat{\mathbf{v}}_1^H(\Omega)\hat{\mathbf{v}}_1(\Omega)}} \tag{6.18}$$

und folglich die PCA-Filterkoeffizienten zu

$$\mathbf{F}_{\mathrm{PCA}}(\Omega) = \frac{1}{\sqrt{M}}\frac{\hat{\mathbf{v}}_1(\Omega)}{||\hat{\mathbf{v}}_1(\Omega)||} \quad \Rightarrow \quad \mathbf{F}_{\mathrm{PCA}}^H\mathbf{F}_{\mathrm{PCA}} = \frac{1}{M}. \tag{6.19}$$

Sieht man den PCA *Beamformer* als "selbstjustierenden" DSB (zumindestens für geringe Nachhallzeiten), so ist die Normierung äquivalent zu der des DSBs in Gl. (3.31): $\mathbf{F}_{\mathrm{DSB}}(\Omega) = \mathbf{d}(\Omega,\theta_t)/M$, mit $||\mathbf{d}(\Omega,\theta_t)|| = \sqrt{M}$. Die einkanalige Nachfilterung bzw. Normalisierung Gl. (6.19) kann bei der Verwendung von Algorithmus 3 (S-Grad-GG) oder Algorithmus 4 (S-Grad-IS) sehr einfach durch die Wahl von $C^2 = 1/M$ ohne zusätzliche Rechenoperationen realisiert werden.

Um die Resultate der Normalisierung des PCA *Beamformers* zu visualisieren, soll das *Beampattern* für alle relevanten Frequenzen und Winkel betrachtet werden. Dazu wurden akustische Sprachdaten nach *Szenario-1* für $M = 5$ Sensoren erzeugt und mit unkorreliertem bzw. diffusem Rauschen überlagert. Das Sprachsignal fällt also aus einer Richtung von 45° bezüglich *Broadside* auf das *Array* ein. Die Filterkoeffizienten sind mit Hilfe von Algorithmus 3 (S-Grad-GG) mit dem Wert $C^2 = 1/M$ und einer Filterlänge von $B = 128$ bestimmt worden.

In Bild 6.1 sind verschiedene Richtcharakteristiken des PCA *Beamformers* in Form einer zweidimensionalen Darstellung von Grauwerten zu sehen. Eine hohe Dämpfung wird durch die Farbe Schwarz und keine Dämpfung durch die Farbe Weiß charakterisiert.

Prinzipiell bildet sich bei der Verwendung des PCA *Beamformers* eine ähnliche Richtcharakteristik wie bei einem *Delay-and-Sum-Beamformer* aus. Zusätzlich zu der konstruktiven Überlagerung der Signalkomponenten welche über die direkte Sichtverbindung auf die Mikrophone einfallen werden allerdings noch frühe Reflexionen berücksichtigt (vgl. Abschnitt 4.5). Auf den exemplarischen Darstellungen der Richtcharakteristik in Bild 6.1 sind folgende Eigenschaften abzulesen:

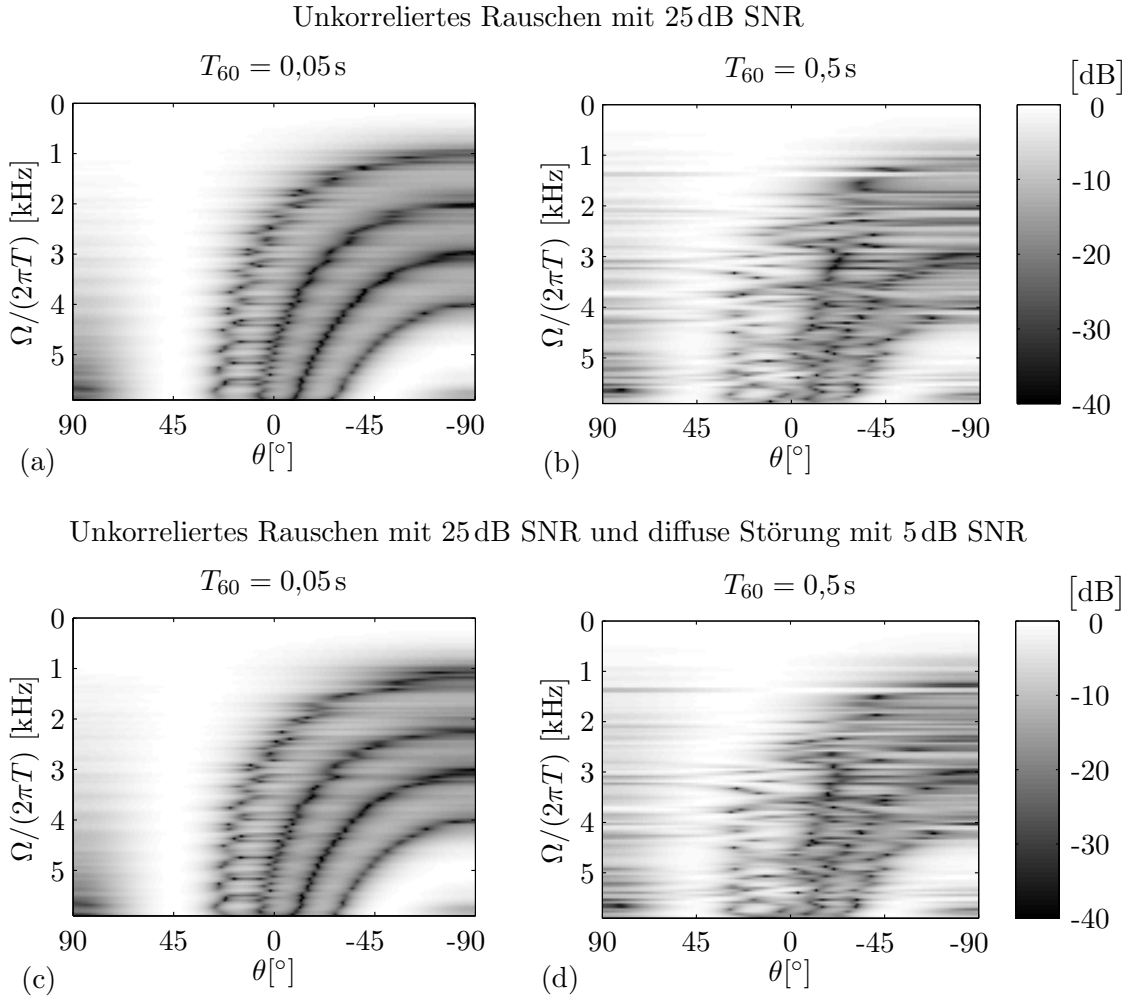

Bild 6.1: Richtcharakteristiken des PCA *Beamformers* für eine Sprecherrichtung von $\theta_s = 45\,^\circ$ und den Nachhallzeiten $T_{60} = 0{,}05\,\mathrm{s}$ sowie $T_{60} = 0{,}5\,\mathrm{s}$. In (a) und (b) mit additivem unkorrelierten Rauschen von 25 dB und in (c) und (d) für zusätzliches diffuses Rauschen von 5 dB SNR.

- Es erfolgt eine automatische Ausrichtung auf die Sprecherrichtung $\theta_s = 45\,^\circ$.
- Für kleine Nachhallzeiten wird für alle Frequenzen die gleiche Dämpfung des Signals von 0 dB aus der Richtung $\theta_s = 45\,^\circ$ erreicht.
- Bei höheren Nachhallzeiten ist nur näherungsweise die gleiche Dämpfung des Signals aus $\theta_s = 45\,^\circ$ aufgrund der mitberücksichtigten Reflexionen erzielbar.
- Diffuses Rauschen hat keinen signifikanten Einfluss auf das PCA *Beamforming*.

6.4.2 GEV Beamforming

Für den GEV *Beamformer* sollen zunächst Richtdiagramme und anschließend erzielbare SNR-Gewinne in Kombination mit dem perzeptuellen Sprachqualitätsmaß PSM präsentiert werden. Diese sind für unterschiedliche geometrische Anordnungen sowie verschiedene Parametereinstellungen untersucht worden. Grundsätzlich wird bei allen Simulationen den Eingangsdaten jeweils weißes, räumlich unkorreliertes Rauschen mit einem SNR von 25 dB hinzugefügt. Des-

weiteren ist die Matrix der Kreuzleistungsdichten der Störung immer mit einem Regularisierungsterm von −40 dB versehen worden.

Beampattern

Im Gegensatz zum PCA *Beamformer* bildet der GEV *Beamformer* bei Vorhandensein einer diffusen Störung im niederfrequenten Bereich eine gänzlich andere Richtcharakteristik aus. Die Hauptkeulen werden dort schmaler, wodurch die aus allen Richtungen einfallende Störung besser unterdrückt werden kann; die Direktivität des *Beamformers* ist somit deutlich ausgeprägter. Dieses Verhalten ist an den in Bild 6.2 dargestellten Richtcharakteristiken für den GEV *Beamformer* ohne und mit nachgeschalteten Normalisierungsverfahren zu erkennen. Die Anzahl der Filterkoeffizienten beträgt $B = 128$ bei einer Verarbeitungsblocklänge von $L = 2B$. Die Koeffizienten wurden mit Hilfe von Algorithmus 5 (A-PM-GG) für das *Szenario-1* mit $M = 5$ Mikrophonen und zusätzlicher Überlagerung von unkorreliertem sowie diffusem Rauschen bestimmt. An dem *Beampattern* für den Fall ohne Normalisierung in Bild 6.2 ist die entstehende Signalverzerrung aufgrund der unterschiedlichen Skalierung von bis zu 15 dB Differenz bei der Einfallsrichtung $\theta_s = 45\,^\circ$ erkennbar. Abhilfe verschaffen hier alle der vorgestellten Nachfilter BAN, BSN und MN. Das recheneffizienteste Verfahren BAN benötigt keinerlei Information über die *Array*-Geometrie und die Sprecherrichtung. Für das BSN-Verfahren ist hingegen der Abstand der Mikrophone zueinander als bekannt vorausgesetzt. Wegen der Berücksichtigung aller Raumrichtungen in der Normalisierung kommt es zu einer leichten Verstärkung des Sprachsignals bei den niedriegen Frequenzanteilen, was daran zu erkennen ist, dass der maximale Wert, gekennzeichnet durch die Farbe Weiß, bei ca. 4 dB liegt. Bei der Maximum-Normalisierung wird wieder der Abstand der Mikrophone zueinander benötigt. Da nun auf den maximalen Wert des *Beampatterns* pro Frequenzkomponente normiert wird, ist hier keinerlei Verstärkung des Signals größer 0 dB zu beobachten.

Für eine gerichtete Störung nach *Szenario-2* bildet sich abhängig von der Nachhallzeit ein ausgeprägtes Minimum im *Beampattern* an der Stelle der Einfallsrichtung des Störsignals bei $\theta_n = -20\,^\circ$ aus. In Bild 6.3 ist das Richtdiagramm für den GEV *Beamformer* ohne Nachfilter für die Nachhallzeit $T_{60} = 0{,}05\,\mathrm{s}$ zu sehen. Da es sich um eine Störquelle mit Tiefpasscharakter handelt, nimmt die Ausprägung des Minimums bei der Richtung $-20\,^\circ$ zu hohen Frequenzen hin ab und läuft in das Minimum der DSB-Richtcharakteristik aus (vgl. Bild 6.1). Im Vergleich zu dem *Beampattern* ohne Normalisierung in Bild 6.2 ist zu erkennen, dass hier die resultierende Sprachverzerrung für das Nutzsignal aus der Richtung $\theta_s = 45\,^\circ$ geringer ausfällt. Die Dämpfung der räumlichen Übertragungsfunktion variiert bei der Sprecherrichtung weniger stark im Vergleich zum Richtdiagramm des diffusen Rauschens Bild 6.2.

Das Verhalten der Nachfilterungsalgorithmen für das gerichtete Rauschen ist in Bild 6.4 an den resultierenden Richtdiagrammen für die Nachhallzeit $T_{60} = 0{,}05\,\mathrm{s}$ in der linken Spalte und $T_{60} = 0{,}5\,\mathrm{s}$ in der rechten Spalte zu sehen. Bei der geringen Nachhallzeit ergibt sich jeweils ein klares Maximum an der Stelle der Sprecherrichtung, wobei hier wieder eine leichte Verstärkung für das BSN-Verfahren von ca. 3 dB auftritt. In den Richtdiagrammen für die hohe Nachhallzeit scheint die Richtcharakteristik etwas zu "verschwimmen". Bei genauerer Betrachtung sind jedoch die beiden Eigenschaften Gl. (6.11) und Gl. (6.12) zu erkennen. Das Maximum des *Beampattern* liegt weiterhin in einer sehr nahen Umgebung um θ_s herum

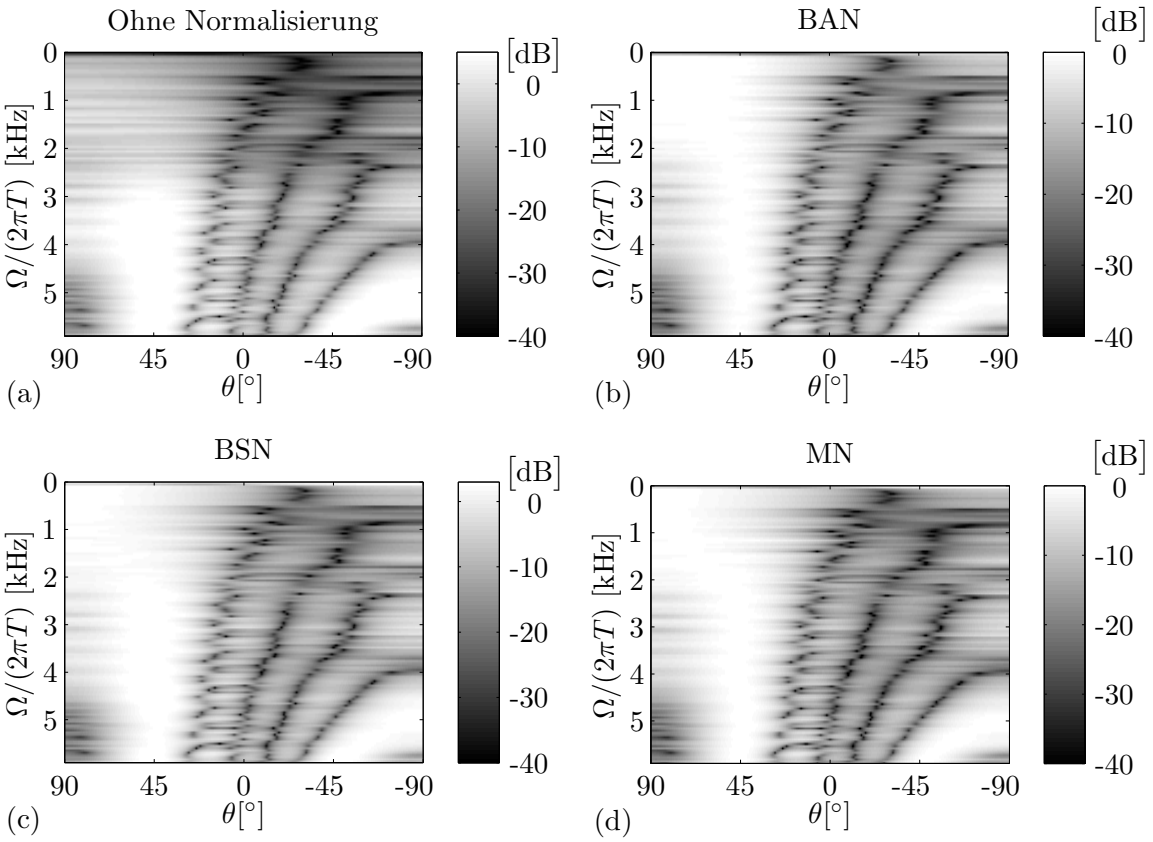

Bild 6.2: Richtcharakteristiken des GEV *Beamformers* ohne und mit unterschiedlichen Normalisierungsverfahren. Die Sprecherrichtung beträgt $\theta_s = 45\,^\circ$, die Nachhallzeit ist $T_{60} = 0{,}05\,\mathrm{s}$ und es wurde unkorreliertes sowie diffuses Rauschen von 25 dB bzw. 5 dB SNR dem Sprachsignal überlagert.

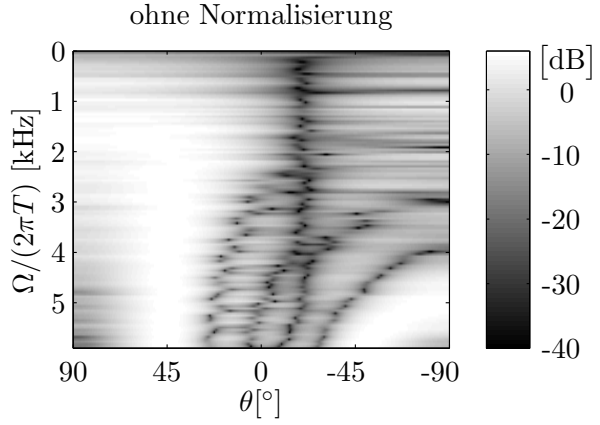

Bild 6.3: Richtcharakteristik des GEV *Beamformers* ohne Nachfilter für die Nachhallzeit vont $T_{60} = 0{,}05\,\mathrm{s}$. Die Sprecherrichtung ist $\theta_s = 45\,^\circ$ und das gerichtete Tiefpassrauschen hat eine Einfallsrichtung von $\theta_n = -20\,^\circ$ bei einem SNR von 5 dB

und an der Stelle θ_n ergibt sich ein ausgeprägtes Minimum. Wie stark sich letztendlich die Nachfilterverfahren auf die akustische Qualität des *Beamformer*-Ausgangs auswirkt, soll im Folgenden ausgewertet werden.

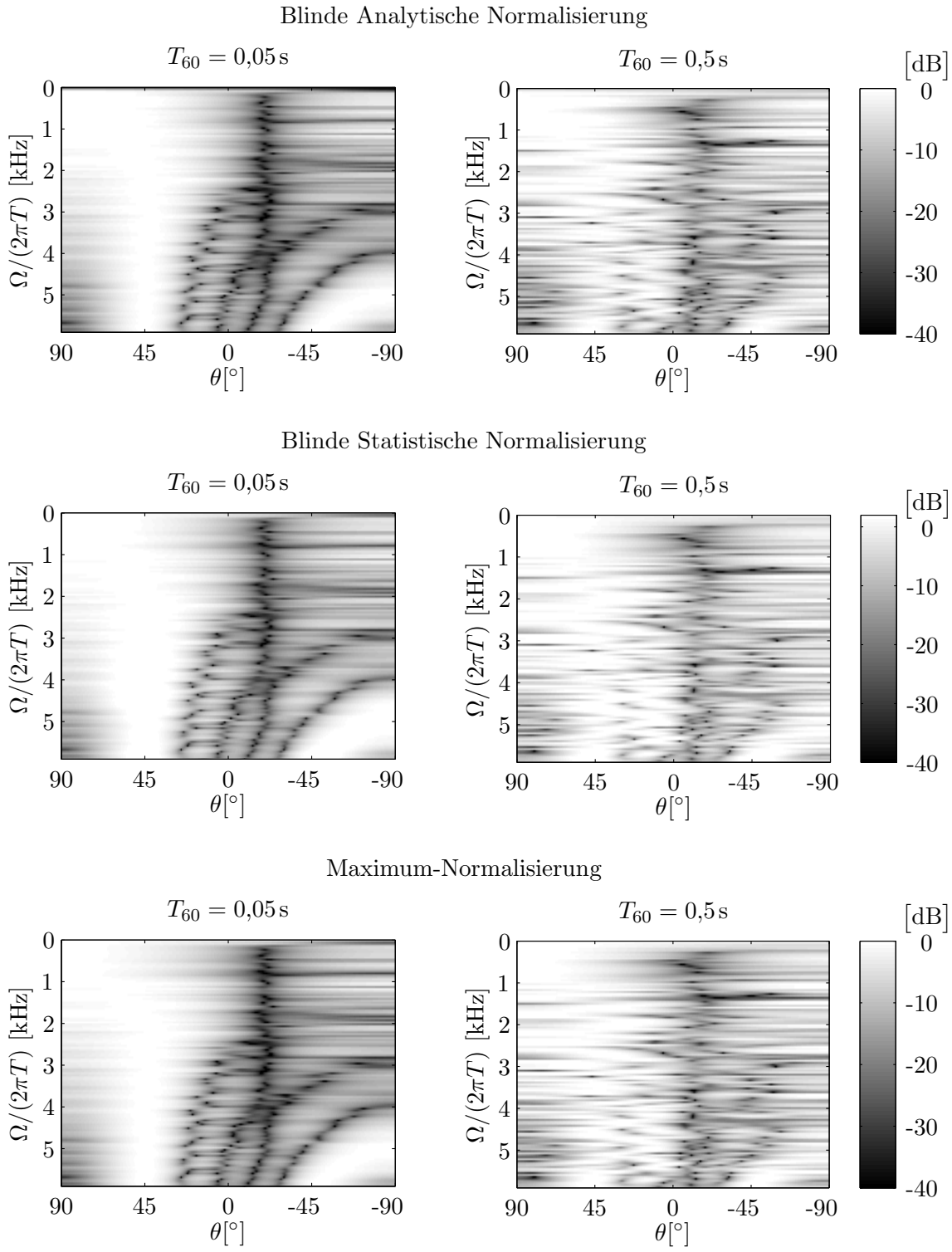

Bild 6.4: Richtcharakteristiken des GEV *Beamformers* mit unterschiedlichen Normalisierungsverfahren für die Nachhallzeiten von $T_{60} = 0{,}05\,\mathrm{s}$ und $T_{60} = 0{,}5\,\mathrm{s}$. Die Sprecherrichtung beträgt $\theta_s = 45^\circ$ und das gerichtete Tiefpassrauschen hat eine Einfallsrichtung von $\theta_n = -20^\circ$.

SNR-Gewinne und PSM-Werte für unterschiedliche geometrische Anordnungen

Für das *Szenario-2* sind die Verläufe des SNR-Gewinns in Bild 6.5 (a) und die Verläufe des perzeptiven Qualitätsmaßes in Bild 6.5 (b) jeweils über der Nachhallzeit aufgetragen. Als

Referenz soll hier der GMVDR *Beamformer* dienen. Für diesen sind die Filterkoeffizienten mit Gl. (4.28) bestimmt worden, wobei die Raumübertragungsfunktion mittels Algorithmus 1 (S-PM-GG) mit einer Blocklänge von $L = 256$ aus den reinen Sprachdaten geschätzt[6] wurde. Diese und die folgenden gemessenen Ergebnisse basieren auf konvergierten Filterkoeffizienten. Zur Ermittlung der PSM-Werte wurden nur die reinen Sprachdaten mit diesen Filterkoeffizienten gefiltert um die Auswirkung der Nachfilteralgorithmen auf die Sprachverzerrung separat ohne zusätzliche Störgeräusche zu analysieren. Das reine, verhallte, mit den GMVDR-Koeffizienten gefilterte Sprachsignal dient also jeweils als Referenzsignal. Und die reinen, verhallten, mit den GEV-Verfahren gefilterten Sprachsignale werden jeweils als Testsignal gegenüber dem Referenzsignal verglichen.

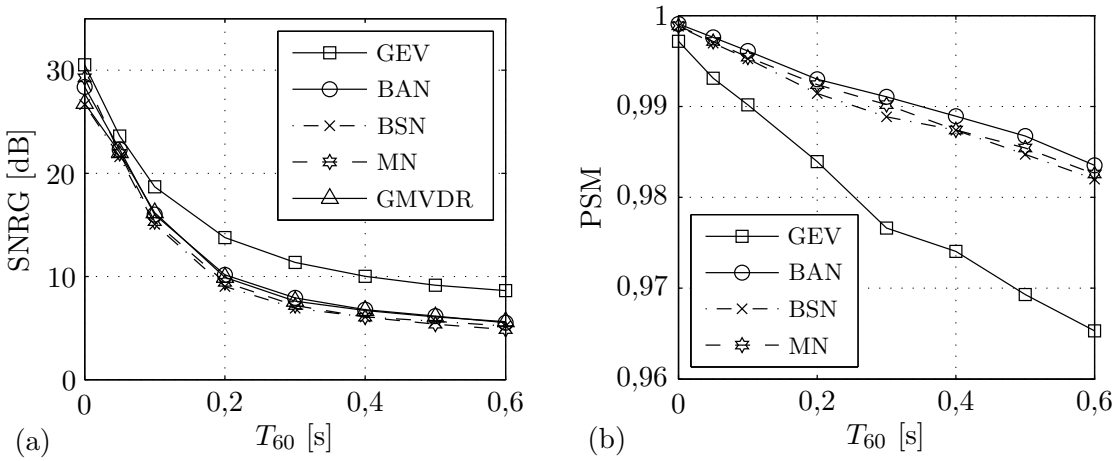

Bild 6.5: SNR-Gewinn in (a) und das perzeptive Qualitätsmaß in (b) für eine Sprecherrichtung von $\theta_s = 45\,^\circ$ und einer Störquelle bei $\theta_n =$ -20 °.

An dem relativ hohen SNR-Gewinn des GEV *Beamformers* ohne Normalisierung (bezeichnet mit GEV) in Bild 6.5 (a) lässt sich erahnen, dass an dessen Ausgang Sprachverzerrungen zu erwarten sind. Genau dieses Verhalten spiegeln in der rechten Darstellung die deutlich geringeren PSM-Werte im Vergleich zum *Beamforming* mit Normalisierungsverfahren (bezeichnet mit BAN, BSN und MN) wieder. Werden die GEV-Filterkoeffizienten mit den beschriebenen Nachfiltern normalisiert, so ergibt sich ein sehr ähnliches Verhalten bezüglich der Störgeräuschreduktion wie für den GMVDR *Beamformer*. Die Sprachqualität lässt sich bei subjektiven Hörtests ebenfalls deutlich dichter den optimal gefilterten Signalen zuordnen, als dies durch das Diagramm in Bild 6.5 (b) ausgedrückt wird. Sie kommen also in der Qualität den Referenzsignalen sehr nahe, wohingegen die Filterung ohne Normalisierung je nach spektraler Zusammensetzung der Störung [7] zur unkontrollierten Verstärkung bzw. Dämpfung einzelner Spektralkomponenten führen kann.

Als nächstes sind in Bild 6.6 die Ergebnisse für das *Szenario-3* dargestellt. Hierbei fällt das Sprachsignal von *Broadside*, aus einer Distanz von 0,8 m, auf das *Array* ein. Es befindet

[6] Da bei dem idealisierten Fall keinerlei Rauschen dem Sprachsignal überlagert ist, kann die BAN bei der Bestimmung der Raumübertragungsfunktion verwendet werden.

[7] Bei der Potenzmethode wird nach jedem Iterationsschritt κ der geschätze Vektor auf die Einheitslänge normiert: $\frac{\hat{\mathbf{v}}_{1,\kappa}}{||\hat{\mathbf{v}}_{1,\kappa}||} = \frac{\zeta}{|\zeta|}\frac{\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}\mathbf{H}}{\sqrt{\mathbf{H}^H\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-2}\mathbf{H}}}$. Im Gegensatz zu den optimalen GMVDR-Filterkoeffizienten ist hier also zu sehen, dass $\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}$ in quadrierter Form im Nenner vorkommt.

sich eine Störquelle in 1,6 m Abstand zum *Array* und bei einer Richtung von $\theta_n = 60\,^\circ$. Für den SNR-Gewinn des GMVDR *Beamformers* und des GEV *Beamformers* mit Filternormalisierung ergeben sich ähnliche Verläufe wie für das *Szenario-2* in Bild 6.5 (a). Bei dem GEV *Beamformer* ohne Nachfilter sieht in Bild 6.6 (a) die Kurve jedoch anders aus: für kleine Nachhallzeiten ergibt sich ein leicht überhöhtes und für hohe Nachhallzeiten ein geringfügig kleineres SNR im Vergleich zu den anderen Verläufen. Das perzeptuelle Maß in Bild 6.6 (b) zeigt jedoch auch hier wie schon vorher deutliche Verzerrungen in der gefilterten Sprache an. Im Gegensatz zum vorherigen Szenario ist nun die Sprachqualität für die Verfahren mit normalisierten Filterkoeffizienten noch etwas angestiegen. Insbesondere ergibt sich für das BAN-Verfahren über alle Nachhallzeiten und alle betrachteten Sprachbeispiele ein minimal homogeneres Klangbild. Generell hat sich bei den Experimenten gezeigt, dass alle Nachfilterverfahren für eine *Broadside*-Ausrichtung die besten Ergebnisse bezüglich der Sprachqualität liefern.

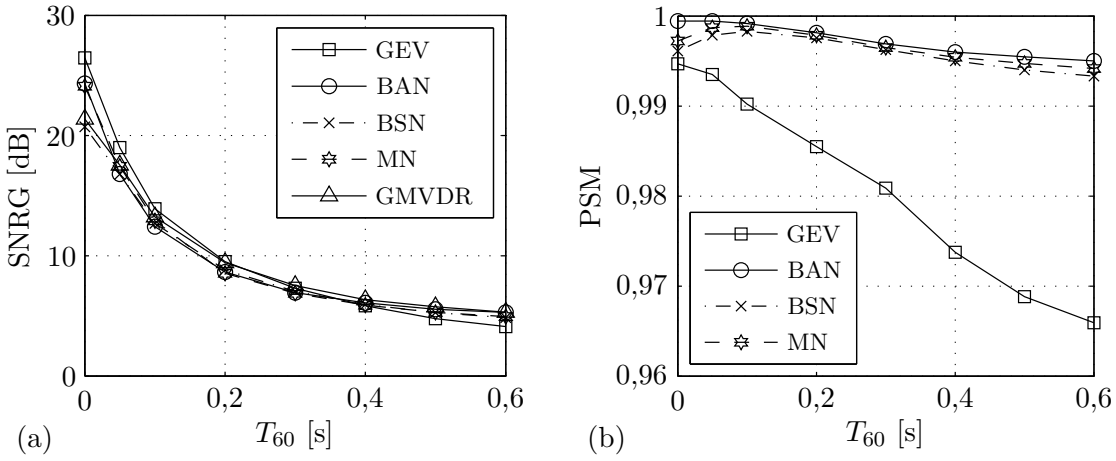

Bild 6.6: SNR-Gewinn in (a) und das perzeptive Qualitätsmaß in (b) für eine Sprecherrichtung von $\theta_s = 0\,^\circ$ und einer Störquelle bei $\theta_n = 60\,^\circ$.

In der Anordnung nach *Szenario-4* fällt das Sprachsignal wieder von *Broadside*, aus einer Distanz von 0,8 m, auf das *Array* ein. Es sind nun zwei Störquellen platziert: eine bei $-20\,^\circ$ und eine bei $60\,^\circ$, jeweils in einem Abstand von 1,6 m zu den Mikrophonen. Bei dieser Anordnung sind nun deutliche Ausprägungen der Sprachverzerrung des GEV *Beamformings* ohne Nachfilterung bei geringen Nachhallzeiten in Bild 6.7 (b) zu beobachten. Dafür liegt der SNR-Gewinn weit über den Werten des SNR-Gewinns des GMVDR *Beamformers*. Dessen Störgeräuschreduktion liegt insgesamt deutlich tiefer im Vergleich zu den anderen Szenarien, da hier eine komplexere Anordnung aus zwei gerichteten Störquellen vorliegt. Die BSN- und MN-Verfahren zeigen bei dieser Anordnung für geringe Nachhallzeiten eine Verfälschung des Sprachsignals durch eine leichte Anhebung der tiefen Frequenzkomponenten. Da bei geringen Nachhallzeiten zwei ausgeprägte Minima entstehen, bildet sich ein recht komplexes *Beampattern* aus. Die Normalisierungsmethoden, basierend rein auf diesem *Beampattern*, zeigen hier nun leichte Schwächen. Hingegen arbeitet die blinde analytische Normalisierung weiterhin sehr zuverlässig und mit durchweg guten Ergebnissen.

Abschließend sollen noch explizit Ergebnisse für den SNR-Gewinn und die resultierende

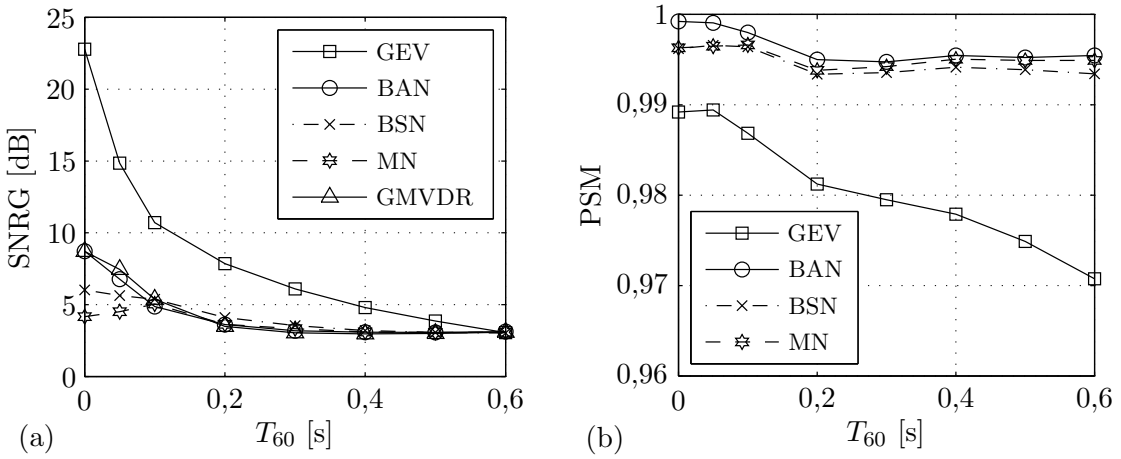

Bild 6.7: SNR-Gewinn in (a) und das perzeptive Qualitätsmaß in (b) für eine Sprecherrichtung von $\theta_s = 0\,^\circ$ und zwei Störquellen: eine bei $-20\,^\circ$ und eine bei $60\,^\circ$.

Sprachqualität für einen Sprecher nach *Szenario-1* in einem diffusen Störschallfeld präsentiert werden. Das SNR am Eingang beträgt dabei wieder 5 dB. Die SNR-Gewinne in Bild 6.8 (a) fallen erwartungsgemäß geringer aus als für die Anordnungen mit gerichteten Störschallquellen. Auffallend sind hier die schlechtesten Werte für die Störgeräuschreduktion bei dem GEV *Beamformer* ohne Nachfilter und die entstehenden Sprachverzerrungen bei niedrigen Nachhallzeiten. Auch bei dieser Anordnung zeigt wieder die BAN-Methode das beste Leistungsverhalten der vorgestellten Nachfilterverfahren.

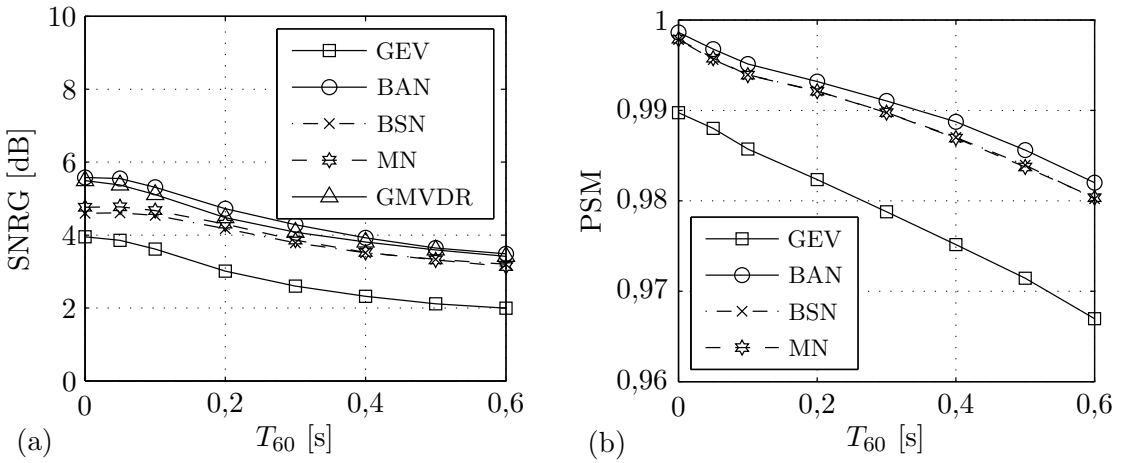

Bild 6.8: SNR-Gewinn in (a) und das perzeptive Qualitätsmaß in (b) für eine Sprecherrichtung von $\theta_s = 45\,^\circ$ und einem diffusen Störschallfeld.

SNR-Gewinne und PSM-Werte für verschiedene Parametereinstellungen

Für das *Szenario-2* wird der Einfluss folgender Parameter exemplarisch untersucht: die Anzahl der Filterkoeffizienten B, das Eingangs-SNR und die Anzahl der Mikrophone M. Dafür wird ausschließlich der GEV *Beamformer* mit BAN-Methode verwendet.

Die Verläufe in Bild 6.9 zeigen die Auswirkung für die Wahl unterschiedlicher Werte von $B \in \{64, 128, 256, 512\}$. Dabei beträgt die Verarbeitungsblocklänge, also die Länge der

Fourier-Transformation wieder jeweils $L = 2B$. Das SNR am Eingang wurde auf 5 dB gesetzt und die Mikrophonanzahl beträgt $M = 5$.

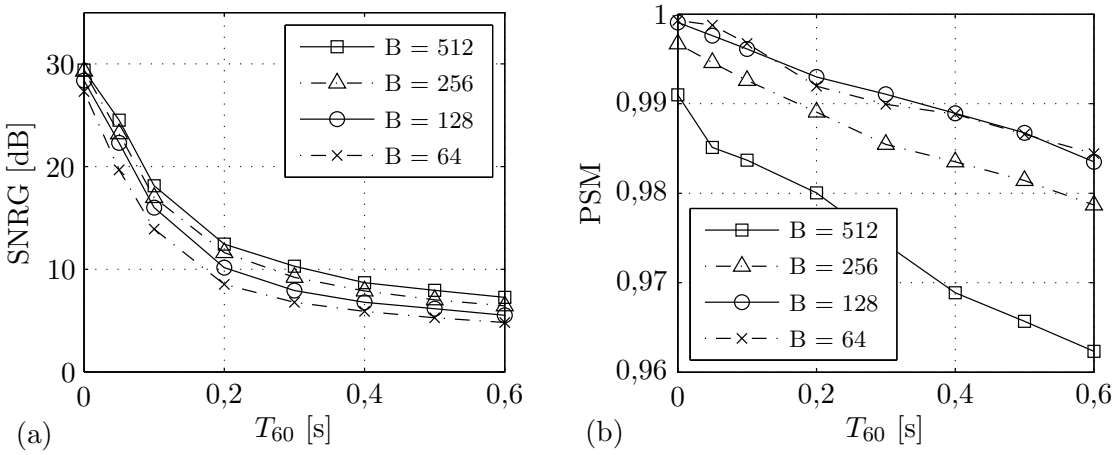

Bild 6.9: SNR-Gewinn in (a) und das perzeptive Qualitätsmaß in (b) für eine variierende Filterlänge B bei dem *Szenario-2*. Das SNR am Eingang beträgt 5 dB und die Anzahl der Mikrophone ist $M = 5$.

An dem SNR-Gewinn in Bild 6.9 (a) ist eine ansteigende Störgeräuschreduktion für größere Filterlängen beobachtbar. Diese beruht prinzipiell auf der Tatsache, dass eine genauere Berechnung der Matrix der Kreuzleistungsdichten der Störung bei steigender Verarbeitungsblocklänge L möglich ist (vgl. Gl. (4.65)). Im Gegensatz dazu wird jedoch die Schätzung der Filterkoeffizienten bei gleichzeitiger Aktivität der Störung mit ansteigender Koeffizientenanzahl ungenauer, was sich an einer stärkeren Verfälschung der spektralen Zusammensetzung im Ausgangssignal bemerkbar macht. Dieses Verhalten ist sehr gut an den fallenden PSM-Werten für steigende Filterlängen in Bild 6.9 (b) zu erkennen. Insgesamt hat sich bei zahlreichen Experimenten eine Filterlänge von $B = 128$ als guter Kompromiss zwischen Störgeräuschreduktion einerseits und Sprachqualität sowie Rechenkomplexität andererseits erwiesen.

Für die Ergebnisse in Bild 6.10 variiert nun das SNR des gerichteten Tiefpassrauschens an den Mikrophonen bei gleichbleibendem SNR des räumlich unkorrelierten weißen Rauschens von 25 dB. Die Anzahl der verwendeten Mikrophone ist $M = 5$ und die Filterlänge beträgt $B = 128$. Bei sehr geringen Nachhallzeiten ist die Steigerung der Störgeräuschreduktion für größere Verhältnisse von räumlich korreliertem zu räumlich unkorreliertem Rauschen ausgeprägter. Dieses Verhalten kann an Gl. (4.65) abgelesen werden und wurde in Bild 4.5 mit expliziten Simulationen dargestellt. Für komplexer werdende Raumimpulsantworten bei wachsenden Nachhallzeiten wird bei steigendem SNR die Schätzung der optimalen Filterkoeffizienten genauer. Dies ist an den leicht höheren PSM-Werten für größere SNR und höhere Nachhallzeiten in Bild 6.10 (b) abzulesen.

Das Verhalten des GEV *Beamformers* mit BAN-Methode ist für eine variierende Anzahl von verwendeten Mikrophonen $M \in \{3, 5, 7, 9\}$ in Bild 6.11 dargestellt. Das SNR am Eingang wurde auf 5 dB gesetzt und die gewählte Filterlänge beträgt $B = 128$. Hier ist nun ein ausgeprägter SNR-Gewinn für steigende Nachhallzeiten bei der Verwendung von zusätzlichen Mikrophonen zu erkennen. Bei geringen Nachhallzeiten ist die Bildung eines Minimums der räumlichen Übertragungsfunktion des *Beamformers* schon mit nur drei Mikrophonen möglich.

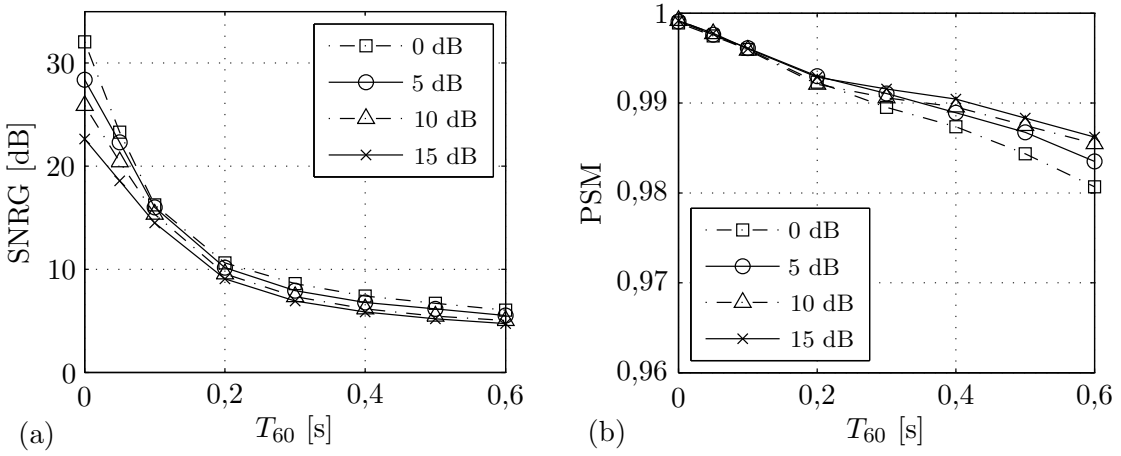

Bild 6.10: SNR-Gewinn in (a) und das perzeptive Qualitätsmaß in (b) für ein variierendes SNR der Mikrophonsignale bei dem *Szenario-2*. Die Filterlänge beträgt $B = 128$ und die Anzahl der verwendeten Mikrophone ist $M = 5$.

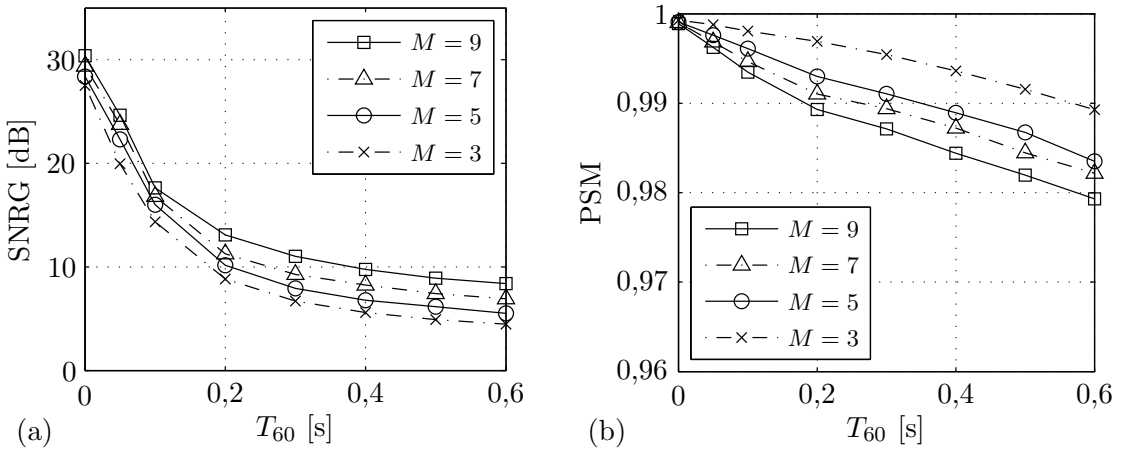

Bild 6.11: SNR-Gewinn in (a) und das perzeptive Qualitätsmaß in (b) für eine variierende Anzahl von verwendeten Mikrophonen M bei dem *Szenario-2*. Die Filterlänge beträgt $B = 128$ und das SNR der Mikrophonsignale beträgt 5 dB.

Daher sind die relativen Unterschiede im SNR Gewinn hier nicht so deutlich. Mit steigender Nachhallzeit wird der Charakter des Störgeräuschfeldes immer diffuser. Damit wird die Näherung für die Störgeräuschunterdrückung mit SNRG $\approx 10\log(M)$ dB immer zutreffender und der relative Unterschied der Verläufe größer. Sehr interessant ist der Unterschied der PSM-Verläufe für die verschiedenen Werte M in Bild 6.11 (b). Hier ist ein umgekehrtes Verhalten im Gegensatz zum SNR-Gewinn erkennbar: mit steigender Anzahl der verwendeten Mikrophone fällt die gemessene Sprachqualität leicht ab. Die Erklärung dafür ist wie folgt: je mehr Sensoren für das *Beamforming* verwendet werden, umso schmaler fällt die sich bildende Hauptkeule pro Frequenzkomponente aus. Bei gleicher Frequenz macht sich aber eine ungenaue Normalisierung umso stärker bemerkbar, je schmaler die entsprechende Hauptkeule ist. Da die blinde analytische Normalisierung nur eine Schätzung darstellt, machen sich also folglich Ungenauigkeiten für eine steigende Anzahl von Mikrophonen stärker bemerkbar.

Die exemplarischen Untersuchungen verschiedener Parameter für das Eigenvektor-*Beamforming* mit Normalisierungsverfahren führen allgemein zu folgenden Aussagen:

- Bei der moderaten Wahl der Filterlänge und der Anzahl der Mikrophone ist ein blindes *Beamforming* mit geringen Sprachverzerrungen bei gleichzeitig guter Störgeräuschreduktion möglich, insbesondere im Fall einer gerichteten Störschallquelle.
- Eine kurze Filterlänge von $B = 128$ ermöglicht ausreichend genaue Schätzungen der Raumübertragungsfunktionen bzw. ihrer Verhältnisse bei simultaner Aktivität der Störschallquellen.
- Eine eher geringe Anzahl von z. B. $M = 5$ Mikrophonen führt zu einem eher kleinen Einfluss von Normalisierungsfehlern und hat zusätzlich den Vorteil einer geringen Rechenkomplexität.

6.5 Zusammenfassung

In diesem Kapitel wurden einkanalige Nachfilter hergeleitet, welche eine Normalisierung der Eigenvektorkoeffizienten pro Frequenzkomponente vornehmen. Das recheneffizienteste Verfahren ist die blinde analytische Normalisierung (BAN), bei der im Wesentlichen eine Matrix-Vektor-Multiplikation notwendig ist. Gleichzeitig weist dieses Verfahren die geringsten Sprachverzerrungen auf. Die beiden weiteren vorgestellten Verfahren nutzen die Struktur des *Beampatterns* aus, welches jedoch relativ aufwendig abgetastet werden muss: die blinde statistische Normalisierung (BSN) normiert die Filterkoeffizienten auf einen mittleren und die Maximum-Normalisierung (MN) auf den maximalen Wert des *Beampatterns*.

Da die Normalisierungsverfahren besser bei einer eher moderaten Wahl für die Filterlänge und die Anzahl der Mikrophone funktionieren, ist die Störgeräuschreduktion für die Anwendung in Räumen mit höheren Nachhallzeiten ebenfalls eher moderat. Der große Vorteil ist jedoch eine schnelle Adaption und somit eine Verfolgung eines sich bewegenden Sprechers. Dies wird aus einer anderen Problemstellung heraus noch in Kapitel 7 demonstriert.

Eine höhere Störgeräuschreduktion in einer aufwendigeren Struktur bei gleichzeitig kaum noch vorhandenen Sprachverzerrungen soll am Schluss dieser Arbeit in Kapitel 8 vorgestellt werden. Dabei ist dann aber von einer eher geringen Sprecherbewegung auszugehen, und außerdem ist eine explizite Bestimmung der Sprecherrichtung notwendig.

Kapitel 7

Sprecherrichtungsbestimmung

Die Sprecherrichtung ist eine wichtige Information für verschiedenste Anwendungen wie z. B. innerhalb einer allgemeinen akustischen Szenenanalyse [SHUW07], in Audio/Video-Konferenzsystemen [WB98, SSR01], zur Sprachsignalsegmentierung und Sprecheridentifikation [SHU06, SHU07], für eine multimodale Mensch-Maschine-Kommunikation [Iri97, LNO00] oder aber zum Laufzeitausgleich in einem *Generalized Sidelobe Canceller* wie er im folgenden Kapitel noch vorgestellt wird.

Nach [DSB01] können bestehende Lokalisationsverfahren grob in drei Kategorien unterteilt werden: Maximierung der Ausgangsleistung eines *Beamformers* durch Steuerung seiner Richtcharakteristik (engl. *Steered Response Power*, SRP), Methoden, welche direkt die Zeitdifferenz der einfallenden Signale mittels Korrelationsverfahren bestimmen (engl. *Time Difference of Arrival*, TDOA) und spektral hochauflösende Verfahren. Für schmalbandige Signale ist in [Sch79] erstmals ein spektral hochauflösendes Verfahren vorgestellt, welches die Bezeichnung MUSIC (*Multiple Signal Classification*) trägt. Dieses findet seither vielfach Anwendung in der Antennentechnik. Dabei ist eine komplette Eigenwert-Dekomposition des Signalraums notwendig, welche insbesondere bei der Erweiterung auf breitbandige Signale wie Sprache sehr rechenintensiv ist. Für Sprachsignale werden daher in der Regel Methoden basierend auf SRP und TDOA eingesetzt [DSB01]. Das Prinzip dieser Verfahren und die Übertragung auf das Eigenvektor-*Beamforming* sollen im Folgenden beschrieben werden. Die Funktionsfähigkeit der neuen Varianten wird durch experimentelle Untersuchungen für verschiedene Schallfelder demonstriert.

7.1 Korrelation der Mikrophonsignale

Die Idee beim TDOA-Verfahren liegt darin, die Zeitverzögerung τ_{il} zwischen zwei Signalen $x_i(t)$ und $x_l(t)$ mittels der Kreuzkorrelation zu bestimmen

$$c_{il}(\tau) = \int_{-\infty}^{\infty} x_i(t) x_l(t+\tau) dt, \quad i,l \in \{1,\ldots,M\} \tag{7.1}$$

$$\tau_{il} = \underset{\tau \in D}{\operatorname{argmax}}\, c_{il}(\tau). \tag{7.2}$$

Stellen die beiden Signale $x_i(t)$ und $x_l(t)$ die zeitkontinuierlichen Mikrophonsignale dar, so ist mit Hilfe der geometrischen Daten der Anordnung eine Umrechnung von τ_{il} in die entsprechende Einfallsrichtung möglich. Dabei ist die Maximumsuche in Gl. (7.2) auf das durch die

geometrische Anordnung bedingte maximal mögliche Intervall D beschränkt. Für die Umsetzung von Gl. (7.1) auf einem digitalen Rechner sind die abgetasteten, zeitdiskreten Signale zu betrachten. Desweiteren ist zwecks Effizienzsteigerung die Berechnung im Frequenzbereich sinnvoll. Zunächst soll hierfür die zeitdiskrete Fourier-Transformation betrachtet werden.

In [KC76] ist eine verallgemeinerte Kreuzkorrelation (engl. *Generalized Cross Correlation*, GCC) vorgestellt worden, die hier definiert wird zu

$$r_{il}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \big(G_i(\Omega) X_i(\Omega)\big) \big(G_l(\Omega) X_l(\Omega)\big)^* e^{j\Omega n} d\Omega, \tag{7.3}$$

wobei die Verschiebung n_{il} zwischen den Signalen aus der Maximumsuche im zu D äquivalenten Intervall N_D hervorgeht

$$n_{il} = \underset{n \in N_D}{\operatorname{argmax}}\, r_{il}(n). \tag{7.4}$$

Die Verallgemeinerung ist auf die beiden spektralen Gewichtungsfunktionen $G_i(\Omega)$ und $G_l(\Omega)$ in Gl. (7.3) zurückzuführen. In [KC76] wurden fünf verschiedene Varianten von Gewichtungsfunktionen beschrieben, wovon sich zwei in praktischen Systemen durchgesetzt haben. Eine basiert auf dem SNR des zu analysierenden Signals und wird als *Maximum-Likelihood*-Gewichtungsfunktion bezeichnet [MA04]. Dabei werden diejenigen Spektralkomponenten in Gl. (7.3) akzentuiert, die wenig Rauschen enthalten. Die häufigste Methode ist jedoch, ausschließlich die Phaseninformation der zu vergleichenden Signale zu nutzen. Diese Phasentransformation (engl. *Phase Transform*, PHAT) ergibt sich durch folgende Gewichtungsfunktionen

$$G_i(\Omega) = \frac{1}{|X_i(\Omega)|}, \qquad G_l(\Omega) = \frac{1}{|X_l(\Omega)|}. \tag{7.5}$$

Die Leistungsfähigkeit der PHAT-GCC wurde in zahlreichen Publikationen gezeigt und auch mit theoretischen Grenzen basierend auf statistischen Modellen der Schallausbreitung verglichen [GRT03]. Zusätzlich zu den in [KC76] aufgeführten Gewichtungsfunktionen existieren natürlich noch weitere, je nach konkreter Anwendung. Da ja die Einfallsrichtung von Sprachsignalen detektiert werden soll, ist z. B. in [Bra99, RYPD05] die Charakteristik von stimmhaften Lauten in der Sprache ausgenutzt worden. Werden z. B. viele Mikrophone verteilt im Raum angeordnet, ist es weiterhin sinnvoll, die jeweiligen Richtungsschätzungen wiederum geeignet gewichtet zu einer Positionsbestimmung zusammenzuführen [MA04, SHUW07].

Betrachtet man nun gemäß der Signalbeschreibung in Abschnitt 3.1 die Einzelkomponenten, aus denen ein Signal in Gl. (7.3) besteht

$$X_i(\Omega) = S_c(\Omega) H_i(\Omega) + N_c(\Omega) A_i(\Omega) + N_{u,i}(\Omega), \tag{7.6}$$

so sind folgende Probleme erkennbar:

- Zur Bestimmung der Einfallsrichtung des Sprachsignals, muss auch der Sprecher aktiv sein ($S_c(\Omega) \neq 0$). Bei einer blockweisen Verarbeitung ist also eine Auswertung für die Signalabschnitte vorzunehmen, in denen auch das Sprachsignal enthalten ist.

- Nach Möglichkeit sollte keine weitere gerichtete Quelle aktiv sein ($N_c(\Omega) = 0$), da sonst auch keine zuverlässige Schätzung der Sprecherrichtung erfolgen kann. Ist dies nicht sicherzustellen, so müssen entweder Verfahren verwendet werden, die mehrere Schallquellen lokalisieren können [DCP01], oder es ist, im Falle von gerichteten Störschallquellen, die hier im weiteren Verlauf vorgestellte Methode einzusetzen.

- Werden die ersten beiden Punkte eingehalten, so ergeben sich noch aufgrund des unkorrelierten Rauschterms Ungenauigkeiten in der Richtungsschätzung.

- Insgesamt hängt die Genauigkeit der Schätzung von der Nachhallzeit und der Komplexität von $\mathbf{H}(\Omega)$ ab. Bei geringen Nachhallzeiten und einer Sichtverbindung zwischen den Mikrophonen und dem Mund des Sprechers sind gute Ergebnisse zu erwarten. Bei höheren Nachhallzeiten und einem abgewendeten Kopf des Sprechers können auch die Richtungen früher Reflexionen fälschlicherweise als direkter Pfad detektiert werden. Abhilfe verschaffen hier große Analysefenster von bis zu mehreren hundert Millisekunden und eine Glättung der instantanen Schätzergebnisse.

Das "ideale Eingangssignal" ist folglich bestimmt durch $X_i(\Omega) = S_c(\Omega)H_i(\Omega)$. Setzt man dies in Gl. (7.3) ein und verwendet die Gewichtungsfunktionen Gl. (7.5), so erhält man

$$r_{il}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{H_i(\Omega)H_l^*(\Omega)}{|H_i(\Omega)||H_l(\Omega)|} e^{j\Omega n} d\Omega. \tag{7.7}$$

Erinnert man sich nun daran, dass die Eigenvektorbestimmung in Kapitel 5 gerade implizit eine Schätzung der Raumübertragungsfunktion bzw. derer Verhältnisse durchführt, so ist es naheliegend, genau diese Schätzungen in Gl. (7.7) zu verwenden. Der Vorteil dabei ist, dass für die adaptive, iterative Eigenvektorbestimmung mehrere Signalblöcke verwendet werden und somit bereits implizit eine gewisse zeitliche Glättung erfolgt. Weiterhin ist auch im Falle von gerichteten Störschallquellen eine relativ gute Bestimmung der Sprecherrichtung möglich.

Bildet also der zu lokalisierende Sprecher die alleinige, bzw. dominante Schallquelle, so kann zunächst der dominante Eigenvektor $\mathbf{v}_1(\Omega) = \zeta(\Omega)\mathbf{H}(\Omega)$ geschätzt werden. Diese Schätzung $\hat{\mathbf{v}}_1(\Omega) = (\hat{v}_{1,1}(\Omega), \ldots, \hat{v}_{1,M}(\Omega))^T$ wird dann äquivalent zu Gl. (7.7) jeweils für die Komponenten $\hat{v}_{1,i}(\Omega)$ und $\hat{v}_{1,l}(\Omega)$ ausgewertet

$$r_{il}^{(\mathrm{PCA})}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\hat{v}_{1,i}(\Omega)\hat{v}_{1,l}^*(\Omega)}{|\hat{v}_{1,i}(\Omega)||\hat{v}_{1,l}(\Omega)|} e^{j\Omega n} d\Omega. \tag{7.8}$$

Die hochgestellte Bezeichnung "(PCA)" in Gl. (7.8) soll darauf hinweisen, dass die ausgewerteten Koeffizienten aus dem speziellen Eigenwertproblem hervorgehen. Wird hingegen das verallgemeinerte Eigenwertproblem unter Berücksichtigung der Matrix $\mathbf{\Phi_{NN}}(\Omega)$ betrachtet, so ist der dominante Eigenvektor $\mathbf{v}_1(\Omega) = \zeta(\Omega)\mathbf{\Phi_{NN}}^{-1}(\Omega)\mathbf{H}(\Omega)$ zu schätzen. Dessen Schätzung $\hat{\mathbf{v}}_1(\Omega)$ ist dann zunächst von links mit $\mathbf{\Phi_{NN}}(\Omega)$ zu multiplizieren, $\tilde{\mathbf{v}}_1(\Omega) = \mathbf{\Phi_{NN}}(\Omega)\hat{\mathbf{v}}_1(\Omega)$, so dass die resultierenden Komponenten von $\tilde{\mathbf{v}}_1(\Omega) = (\tilde{v}_{1,1}(\Omega), \ldots, \tilde{v}_{1,M}(\Omega))^T$ für die Auswertung hergenommen werden können

$$r_{il}^{(\mathrm{GEV})}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\tilde{v}_{1,i}(\Omega)\tilde{v}_{1,l}^*(\Omega)}{|\tilde{v}_{1,i}(\Omega)||\tilde{v}_{1,l}(\Omega)|} e^{j\Omega n} d\Omega. \tag{7.9}$$

Nun weist der Index "(GEV)" in Gl. (7.9) auf die vorherige Auswertung des verallgemeinerten Eigenwertproblems hin. Durch die Berücksichtigung von $\mathbf{\Phi_{NN}}(\Omega)$ ist also auch eine Bestimmung der Sprecherrichtung möglich, obwohl weitere, gerichtete Schallquellen vorhanden sind. Deren Statistik muss allerdings in $\mathbf{\Phi_{NN}}(\Omega)$ erfasst worden sein.

Die geschätzte Sprecherrichtung folgt für die PCA- und GEV-basierte Kreuzkorrelation aus der gleichen Vorschrift zur Maximumsuche wie in Gl. (7.4).

7.2 Abtastung der Richtcharakteristik

Bereits in [BS73, HT73] ist das Prinzip beschrieben, einen *Beamformer* in verschiedene Richtungen zu steuern und nach Maxima in der Ausgangsleistung zu suchen. Die Richtung korrespondierend zu dem absoluten Maximum kann dann als die Einfallsrichtung der dominanten Quelle kategorisiert werden.

Die Ausgangsleistung eines gesteuerten *Filter-and-Sum-Beamformers*, abhängig von der betrachteten Einfallsrichtung θ, kann geschrieben werden als

$$P(\theta) = \int_{-\pi}^{\pi} \left| \sum_{i=1}^{M} G_i(\Omega) X_i(\Omega) e^{j\Omega n_i(\theta)} \right|^2 d\Omega \tag{7.10}$$

wobei $n_i(\theta)$ in Gl. (7.10) die richtungsabhängige Verschiebung am i-ten Mikrophon gegenüber einer Referenz, z. B. $n_1(\theta) = 0$ beschreibt und $G_i(\Omega)$ die spektrale Gewichtung des i-ten Signalpfades. Die Schätzung der Sprecherrichtung $\hat{\theta}_s$ folgt aus der Maximumsuche

$$\hat{\theta}_s = \underset{\theta}{\operatorname{argmax}}\, P(\theta). \tag{7.11}$$

Man erkennt, dass die Vektorschreibweise der komplex konjugierten Exponentialterme in Gl. (7.10) gerade den *Steering Vektor* $\mathbf{d}(\Omega, \theta)$ ergibt. Setzt man eine Gleichgewichtung von $G_i(\Omega) = 1/M \ \forall i$ an, so resultiert der *Uniformly Weighted Beamformer* aus Abschnitt 3.3, welcher in dieser Arbeit auch synonym als DSB bezeichnet wird. Die einfachste Realisierung der SRP-Methode mittels DSB-Anordnung ist folglich gegeben durch die Maximumsuche[1] in

$$P^{(\mathrm{DSB})}(\theta) = \frac{1}{M} \int_{-\pi}^{\pi} \left| \mathbf{d}^H(\Omega, \theta) \mathbf{X}(\Omega) \right|^2 d\Omega. \tag{7.12}$$

Wird das allgemeine Signal aus Gl. (7.6) in Gl. (7.12) eingesetzt, ergeben sich ähnliche Probleme, wie sie im Abschnitt 7.1 bereits aufgezeigt wurden. Gemäß [DSB01] ist die SRP-Methode im Vergleich zu dem TDOA-Verfahren weniger robust und weist deutlich mehr Nebenmaxima auf.

Betrachtet man allerdings äquivalent zum vorherigen Abschnitt nur die Phase der Eingangssignale durch die entsprechende Wahl von $G_i(\Omega) = M/||\mathbf{X}(\Omega)|| \ \forall i$, was insgesamt einer Normierung auf die mittlere Leistung[2] entspricht, und setzt wieder ausschließlich nur das mehrkanalige reine Sprachsignal in Gl. (7.10) ein, so ergibt sich

$$P(\theta) = M \int_{-\pi}^{\pi} \left| \mathbf{d}^H(\Omega, \theta) \frac{\mathbf{H}(\Omega)}{||\mathbf{H}(\Omega)||} \right|^2 d\Omega, \tag{7.13}$$

was aber gerade das *Powerpattern* der Raumübertragungsfunktion ausgewertet für die Richtung θ darstellt (vgl. Abschnitt 3.3). In Gl. (7.13) läßt sich wieder der Vektor der Raumübertragungsfunktion durch den geschätzten dominanten Eigenvektor ersetzen. Mit dem Koeffizientenvektor $\hat{\mathbf{v}}_1(\Omega)$, resultierend aus dem speziellen Eigenwertproblem, lässt sich dann schreiben

$$P^{(\mathrm{PCA})}(\theta) = M \int_{-\pi}^{\pi} \left| \mathbf{d}^H(\Omega, \theta) \frac{\hat{\mathbf{v}}_1(\Omega)}{||\hat{\mathbf{v}}_1(\Omega)||} \right|^2 d\Omega. \tag{7.14}$$

[1] Für die Maximumsuche in Gl. (7.12) ist der Faktor $1/M$ unerheblich.
[2] Für die Maximumsuche in Gl. (7.13) ist der Faktor M unerheblich.

Die Interpretation von Gl. (7.14) ist also, dass die Richtcharakteristik des PCA *Beamformers* abgetastet wird, und der Wert von θ, für den sich das Maximum dieser Abtastung ergibt, gerade die Schätzung der Sprecherrichtung darstellt. Dieses Vorgehen deckt sich mit den Erkenntnissen aus dem Abschnitt 6, insbesondere bei der Betrachtung der Richtdiagramme in z. B. Bild 6.1 oder Bild 6.3.

Im Falle von gerichteten Störschallquellen ist die Schätzung des dominanten, generalisierten Eigenvektors $\hat{\mathbf{v}}_1(\Omega)$ zunächst wieder von links mit $\mathbf{\Phi_{NN}}(\Omega)$ zu multiplizieren, $\tilde{\mathbf{v}}_1(\Omega) = \mathbf{\Phi_{NN}}(\Omega)\hat{\mathbf{v}}_1(\Omega)$, und $\tilde{\mathbf{v}}_1(\Omega)$ kann dann für eine zu Gl. (7.13) bzw. Gl. (7.14) äquivalente Form genutzt werden

$$P^{(\mathrm{GEV})}(\theta) = M \int_{-\pi}^{\pi} \left| \mathbf{d}^H(\Omega,\theta) \frac{\tilde{\mathbf{v}}_1(\Omega)}{||\tilde{\mathbf{v}}_1(\Omega)||} \right|^2 d\Omega. \tag{7.15}$$

7.3 Implementierungsaspekte und Experimente

Zunächst soll auf Implementierungsaspekte der Eigenvektor-basierten Korrelationsmethode bzw. Abtastung der Richtcharakteristik eingegangen werden. Der erste wesentliche Punkt dabei ist die diskrete Verarbeitung der einzelnen Spektralkomponenten Ω_k im Frequenzbereich. Eng damit verknüpft ist die blockweise Betrachtung der Signale mit dem Blockindex m und die blockweise Iteration der Eigenvektoren $\hat{\mathbf{v}}_{1,m}$. Je nach Ansatz – spezielles oder allgemeines Eigenwertproblem – ergeben sich unterschiedliche Vektoren, die zur kompakteren Schreibweise wie folgt zugewiesen werden sollen

$$\mathbf{F}_m(\Omega_k) = \begin{cases} \dfrac{\hat{\mathbf{v}}_{1,m}(\Omega_k)}{||\hat{\mathbf{v}}_{1,m}(\Omega_k)||} & \text{für PCA-Filterkoeffizienten} \\ \dfrac{\hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k)\hat{\mathbf{v}}_{1,m}(\Omega_k)}{||\hat{\mathbf{\Phi}}_{\mathbf{NN}}(\Omega_k)\hat{\mathbf{v}}_{1,m}(\Omega_k)||} & \text{für GEV-Filterkoeffizienten.} \end{cases} \tag{7.16}$$

Dies führt für jeden Verarbeitungsblock m zu

$$r_{il,m}(n) = \frac{1}{L} \sum_{k=0}^{L-1} \frac{F_{m,i}(\Omega_k) F_{m,l}^*(\Omega_k)}{|F_{m,i}(\Omega_k)||F_{m,l}(\Omega_k)|} e^{j\Omega_k n} \tag{7.17}$$

$$n_{il,m} = \underset{n \in N_D}{\operatorname{argmax}}\, r_{il,m}(n), \tag{7.18}$$

wobei Gl. (7.17) effizient mit der schnellen Fourier-Transformation berechnet werden kann. Nun soll der Einfachheit halber die Fernfeld-Näherung zwecks einfacher Berechnung des Einfallswinkels verwendet werden. Weiterhin sind die Mikrophone linear und äquidistant mit dem Abstand d zueinander angeordnet. Äquivalent zu Gl. (3.33) kann dann mit Gl. (7.18) die Schätzung der Sprecherrichtung für das Mikrophonpaar (i,l) angegeben werden zu

$$\theta_{il,m} = \arcsin\left(\frac{c \cdot n_{il,m}}{f_{Ab} \cdot d \cdot (i-l)} \right), \quad i \neq l. \tag{7.19}$$

Unter Verwendung aller Mikrophonpaare – ohne Permutation – ergibt sich schließlich für die Sprecherrichtung θ_s die gemittelte Schätzung pro Verarbeitungsblock

$$\hat{\theta}_{s,m} = \frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{l=i+1}^{M} \theta_{il,m}. \tag{7.20}$$

Bei dieser Vorgehensweise entsteht ein gewisses Problem bezüglich der Auflösung, die mit Gl. (7.19) erreicht werden kann. Denn bei der hier betrachteten Anwendung ist der Abstand zwischen benachbarten Mikrophonen relativ klein: $d = 0{,}04\,\mathrm{m}$. Nimmt man bei einer Abtastfrequenz von $f_{Ab} = 12\,\mathrm{kHz}$ beispielsweise folgende Verschiebung $n_{il,m} = i - l$ an, also gerade einen Abtastwert zwischen zwei benachbarten Mikrophonen, so ergibt sich für das Paar $(1,2)$ ungefähr der Winkel $45\,^\circ$, für das Paar $(1,3)$ $21\,^\circ$, für das Paar $(1,4)$ $14\,^\circ$ und für das Paar $(1,5)$ $11\,^\circ$. Diese Auflösung ist jedoch deutlich zu gering. Daher ist es sinnvoll eine Interpolation von $r_{il,m}(n)$ in Gl. (7.17) um die Stelle $r_{il,m}(n_{il,m})$ herum durchzuführen. Es wurde ein Interpolationsfilter mit MATLAB nach [IEE79] entworfen und in die Software zur Bestimmung der Sprecherrichtung derart eingebunden, so dass die Anzahl der interpolierten Werte zwischen den Stützstellen variabel eingestellt werden kann. Benutzt man z. B. 16 interpolierte Werte, kann bereits mit zwei benachbarten Mikrophonen eine Einfallsrichtung von $\pm 2{,}5\,^\circ$ detektiert werden.

Für die Methode der Abtastung der Richtcharakteristik ist eine Interpolation nicht notwendig, da der *Steering Vector* für beliebige Winkel direkt berechnet werden kann. Bei $2N+1$ äquidistanten Winkeln

$$\theta_\nu = \frac{\pi}{2N}\nu, \qquad \nu = -N, \ldots, N \tag{7.21}$$

ist mit z. B. $N = 45$ eine ausreichende Auflösung von $2\,^\circ$ eingestellt. Die resultierenden $M \cdot (2N+1)$ Exponentialterme pro Frequenzkomponente im *Steering Vector* können *a priori* berechnet werden, so dass letztlich

$$P(\theta_\nu) = \sum_{k=K_u}^{K_o} \left| \mathbf{d}^H(\Omega_k, \theta_\nu) \mathbf{F}_m(\Omega_k) \right|^2 \tag{7.22}$$

auszuwerten ist. In Gl. (7.22) ist durch die Angabe einer unteren Schranke K_u und einer oberen Schranke K_o mit $0 \leq K_u < K_o \leq L-1$, die Auswahl einer Menge von Spektralkomponenten möglich. Die geschätzte Sprecherrichtung folgt wieder aus einer Maximumsuche

$$\hat{\theta}_s = \underset{\theta_\nu}{\operatorname{argmax}}\, P(\theta_\nu). \tag{7.23}$$

Simulationen

Die Funktionsfähigkeit der vorgestellten Lokalisationsalgorithmen basierend auf der Korrelation der geschätzten Raumübertragungsfunktionen bzw. der Abtastung ihrer Richtcharakteristik soll anhand von anschaulichen Beispielen exemplarisch gezeigt werden. Dazu wurde eine Quelle im Wechsel an zwei Positionen platziert, welche jeweils einen Abstand von $0{,}8\,\mathrm{m}$ zum Mittelpunkt der Mikrophongruppe hatte. Die beiden Einfallsrichtungen des akustischen Signals waren $-45\,^\circ$ und $0\,^\circ$. So wurde ein mehrkanaliges Signal zu einer Datei bestehend aus drei Teilsequenzen zusammengefasst: eine Sprachäußerung bei $-45\,^\circ$, anschließend bei $0\,^\circ$ und wieder eine Äußerung bei $-45\,^\circ$. Dem mehrkanaligen, reinen Sprachsignal wurde jeweils unkorreliertes weißes Rauschen mit einem SNR von $25\,\mathrm{dB}$ hinzuaddiert und wahlweise diffuses bzw. gerichtetes Rauschen mit einem SNR von $5\,\mathrm{dB}$ überlagert.

In allen Fällen ist der Algorithmus 6 (A-PM-EG) zur Bestimmung des verallgemeinerten dominanten Eigenvektors verwendet worden. Die Konstante für die exponentielle Glättung ist zu $\alpha = 0{,}96$ gewählt und die Anzahl der berechneten Koeffizienten beträgt 128. Die so iterativ bestimmten GEV-Filterkoeffizienten werden für jeden Eingangsblock mit Gl. (7.16) in Gl.

(7.17) bzw. Gl. (7.22) ausgewertet, so dass sich einerseits mittels Gl. (7.18), Gl. (7.19) und Gl. (7.20), sowie andererseits mittels Gl. (7.23) die blockabhängigen Schätzungen für die Sprecherrichtung ergeben. Für die Korrelationsmethode sind 16 interpolierte Werte zwischen den Stützstellen um das Maximum herum benutzt worden. Für die Abtastung der Richtcharakteristik soll hier eine Winkelauflösung von einem Grad und ein ausgewerteter Frequenzbereich von 500 Hz bis 5500 Hz verwendet werden.

Die Ergebnisse für die Bestimmung der Sprecherrichtung sind in Bild 7.1 über der Zeit aufgetragen. Das Verfahren mittels der Kreuzkorrelationen der geschätzten Raumübertragungsfunktionen ist mit "XK" bezeichnet und die Abtastung der Richtcharakteristik mit "AR". Zusätzlich ist in Bild 7.1 die tatsächliche Einfallsrichtung zu sehen, wobei diese nur für die drei Zeitabschnitte dargestellt ist, in denen auch Sprachaktivität vorliegt. Daher ist die Bezeichnung "S/W" gewählt worden ("S" für Sprachaktivität und "W" für wahrer Winkel).

An den Ergebnissen für die geringe Nachhallzeit in der linken Spalte von Bild 7.1 ist nun eine Eigenschaft besonders auffällig, und zwar die scheinbare Unabhängigkeit von dem vorliegenden Störschallfeld. In der Tat ist aufgrund der expliziten Berücksichtigung der Kreuzleistungsdichten der Störung bei der Eigenvektorbestimmung eine gute Schätzung für die Einfallsrichtung des Nutzsignals in unterschiedlichsten Anordnungen beobachtet worden. Dies gilt bei kleinen Nachhallzeiten für beide Lokalisationsverfahren. Bei mittleren und höheren Nachhallzeiten weist die Methode durch Abtastung der Richtcharakteristik die genaueren Ergebnisse auf. In der rechten Spalte von Bild 7.1 ist deutlich zu erkennen, dass das Korrelationsverfahren eine ungenauere Schätzung liefert. Hier führt die Kombination aus signifikanten frühen Reflexionen in den Raumimpulsantworten und die Interpolation der Korrelationsergebnisse zu Schätzfehlern, was aufgrund der Anordnung und den damit verbundenen ausgeprägteren Reflexionen bei der Richtung -45° deutlich zu erkennen ist. Es sei dennoch angemerkt, dass bei solch stark gestörten Sprachsignalen, wie sie hier zugrundeliegen, eine Ungenauigkeit von ±5° als sehr gering einzustufen ist.

Die guten Ergebnisse bei der Abtastung der Richtcharakteristik wurden durch einen hohen Rechenaufwand aufgrund der zahlreichen komplexen Multiplikationen in Gl. (7.22) erkauft. Die Berechnungsdauer liegt um ein Vielfaches über der Dauer zur Schätzung der Sprecherrichtung mittels der Korrelationsmethode. Hier verhilft jedoch ein einfacher Trick zu einer deutlichen Komplexitätsreduzierung. Da bei den gewählten Parametern ca. alle 10 ms Gl. (7.22) ausgewertet wird, können einerseits sehr schnell Änderungen der Sprecherrichtung erfasst werden, wie an den Verläufen in Bild 7.1 zu sehen ist. Für eine praktische Anwendung scheint dies jedoch nicht in dem Maße notwendig zu sein. Daher kann eine Berechnung der Werte $|\mathbf{d}^H(\Omega_k,\theta_\nu)\mathbf{F}_m(\Omega_k)|^2$ zwar für alle Winkel θ_ν aber für eine bestimmte Untermenge an Frequenzkomponenten $k = K_u, K_u+\Delta, K_u+2\Delta, ...$ im Abstand Δ für *einen* Verarbeitungsblock m erfolgen, welche dann für *weitere* Verarbeitungsblöcke als konstant erachtet werden. Für den *nächsten* Verarbeitungsblock $m+1$ erfolgt die Aktualisierung von $|\mathbf{d}^H(\Omega_k,\theta_\nu)\mathbf{F}_{m+1}(\Omega_k)|^2$ für die Spektralkomponenten $k = K_u+1, K_u+\Delta+1, K_u+2\Delta+1, ...$, im übernächsten Block $m+2$ für die Komponenten $k = K_u+2, K_u+\Delta+2, K_u+2\Delta+2, ...$ usw., wodurch der Berechnungsaufwand ungefähr noch $1/\Delta$ des ursprünglichen Aufwands beträgt. Weiterhin sollte zur Komplexitätsreduktion eine gröbere Winkelauflösung von z. B. 3° gewählt werden.

In Bild 7.2 sind die Verläufe für die Lokalisationsmethode durch Abtastung der Richtcharakteristik für die Anordnung mit der gerichteten Störschallquelle dargestellt. Es wurde eine Winkelauflösung von 3° durch die Wahl von $N = 30$ in Gl. (7.21) eingestellt und einerseits $\Delta = 1$ sowie andererseits $\Delta = 30$ gewählt.

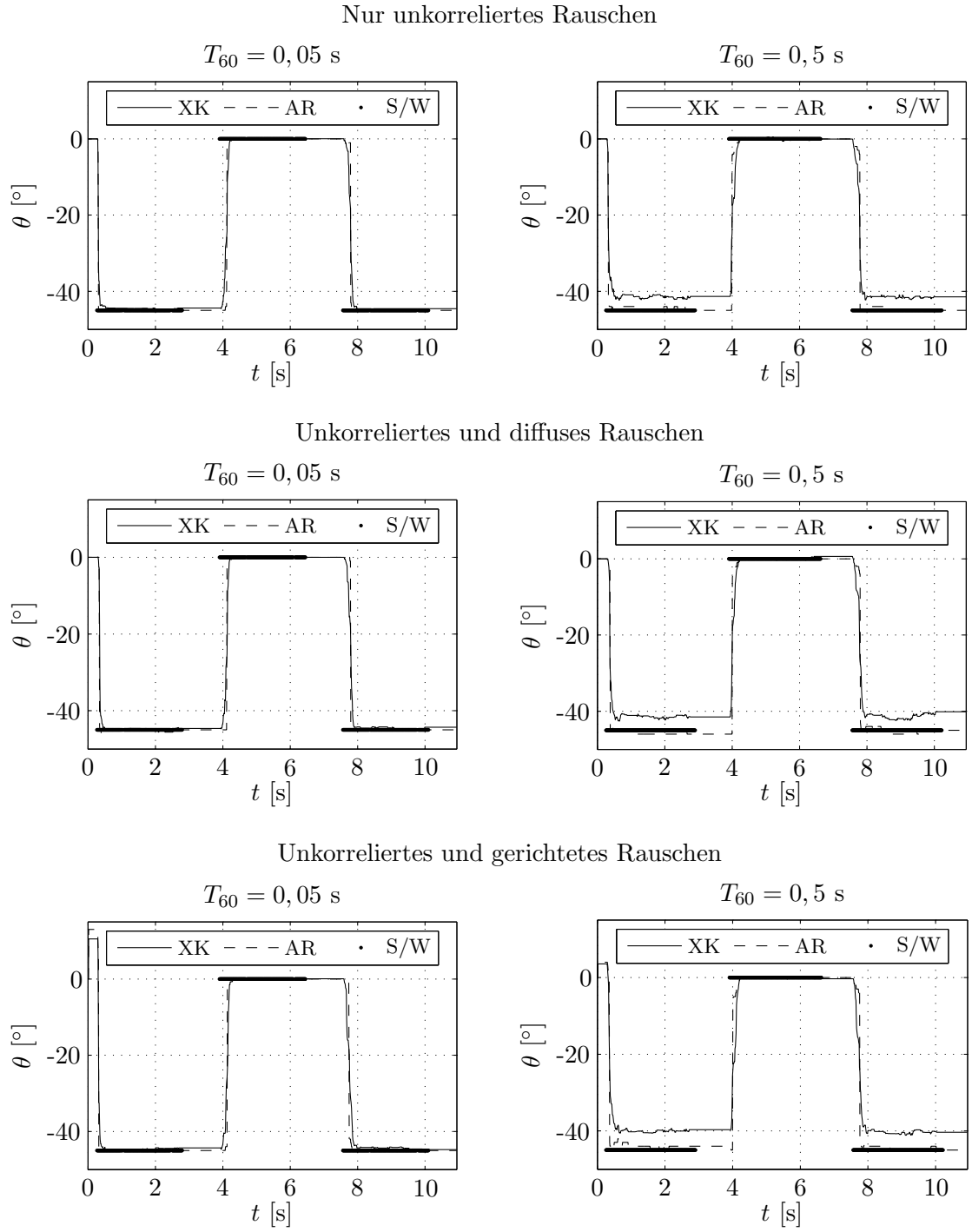

Bild 7.1: Sprecherrichtungsbestimmung mittels Kreuzkorrelationen der geschätzten Raumübertragungsfunktionen "XK" und der Abtastung der Richtcharakteristik "AR". Der tatsächliche Winkel ist mit "W/S" dargestellt und nur für Zeiten mit Sprachaktivität eingetragen.

Die Lokalisiationsergebnisse in Bild 7.2 zeigen zum einen für einige Zeitpunkte Sprünge in der Richtungsschätzung durch die gröbere Winkelauflösung. Zum anderen ist für den Fall der Aktualisierung lediglich jede 30. Spektralkomponente pro Verarbeitungsblock bei der Abtastung der Richtcharakteristik durch die Wahl von $\Delta = 30$ eine sehr geringe Verzögerung in der Nachführung der Sprecherrichtung zu erkennen. Aufgrund der enormen Reduzierung des Berechungsaufwands sind diese beiden Effekte jedoch tolerierbar. Insbesondere, da die

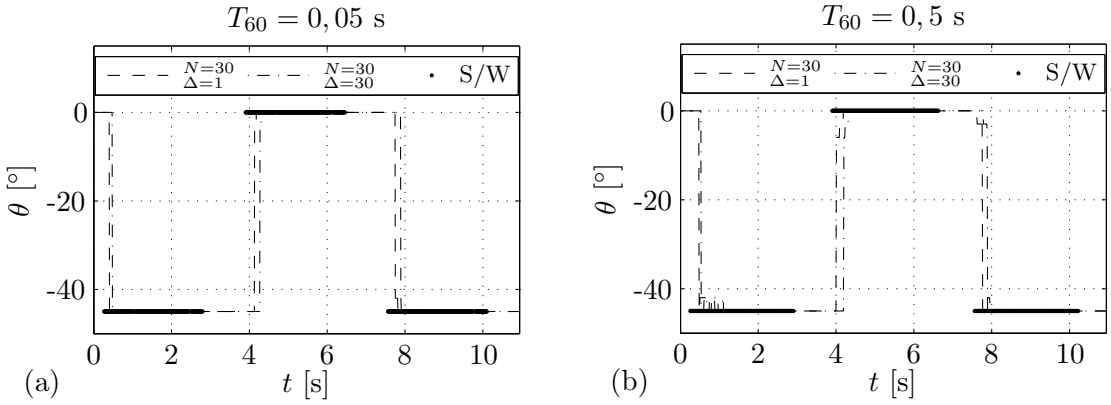

Bild 7.2: Sprecherrichtungsbestimmung mittels der Abtastung der Richtcharakteristik für die Anordnung mit der gerichteten Störschallquelle. Die Winkelauflösung beträgt 3 ° ($N = 30$) und pro Verarbeitungsblock wird einerseits jede Spektralkomponente aktualisiert ($\Delta = 1$) sowie andererseits nur jede 30. Spektralkomponente ($\Delta = 30$).

Verzögerung einer Nachführung der Richtungsschätzung maßgeblich durch das Anzeigen von Sprachaktivität durch die Sprache/Pause-Detektion abhängt.

Zustandsbasierte Nachfilterung

Die in Bild 7.1 und Bild 7.2 gezeigten Ergebnisse stellen instantane Schätzungen pro Verarbeitungsblock dar. Grundsätzlich können diese noch durch z. B. eine Median-Filterung oder eine exponentielle Glättung nachgefiltert werden, um ein robusteres Verhalten gegenüber geringen Positionsänderungen des Sprechers zu erhalten. Für ein Szenario, in dem eine sich kontinuierlich bewegende Schallquelle verfolgt werden soll, können auch aufwendigere Algorithmen zur Weiterverarbeitung genutzt werden. Dabei ist es möglich, eine instantane Positionsschätzung dadurch zu verbessern, indem rekursiv alle bisherigen Beobachtungen durch ein Zustandsmodell in die Schätzung mit einfließen. Ein Zustand enthält dabei die Positions- und Geschwindigkeitsinformation. Dafür ist einerseits ein Messmodell für die Beobachtungen und andererseits ein Bewegungsmodell[3] zur Nachbildung der Bewegungseigenschaften notwendig. Handelt es sich bei den Modellen um lineare Systeme, so kann ein Kalman Filter als stochastischer Zustandsschätzer zur Verfolgung der Sprecherbewegung genutzt werden. Dabei wird jedoch nur die instantane (linearisierte) Positionsschätzung als Beobachtung verwendet. Wählt man z. B. das SRP-Verfahren als Messung für die Wahrscheinlichkeit einer hypothetisierten Sprecherrichtung, so ist es möglich, den durch die Linearisierung entstehenden Informationsverlust zu vermeiden, und jede ausgewertete Richtung wird als Beobachtung herangezogen. Dadurch wird rekursiv die gerade aktuelle, aber unbekannte Wahrscheinlichkeitsdichte auf dem Zustandsraum geschätzt, um daraus den wahrscheinlichsten Systemzustand zu bestimmen. Hierfür wird eine Wolke so genannter Partikel erzeugt, die Paare aus einem Gewicht und einem Punkt im Zustandsraum sind, und als Ganzes die Wahrscheinlichkeitsdichte modellieren. Diese Variante der stochastischen Verfahren zur Zustandsschätzung wird sequenzielle Monte-Carlo-Methode oder aber auch Partikel-Filterung genannt [DFG01, RAG04].

Zur Verfolgung einer Sprecherposition wurde eine Partikel-Filterung erstmals in [VB01] vorgestellt, wobei die Gewichtung aus einer Kreuzkorrelation der Mikrophonsignale – also

[3] In [VB01, WW02, WHUP04] sind Mess- und Bewegungsmodelle für die Problemstellung der Sprecherverfolgung zu finden.

TDOA-Verfahren – berechnet wurde. In [WW02, LWW03, WLW03] kamen robustere Varianten zur Gewichsbestimmung mittels eines gesteuerten DSBs – also SRP-Verfahren – zum Einsatz. Eine Variante der hier vorgestellten Abtastung der Richtcharakteristik von PCA-*Beamformer*-Koeffizienten wurde schließlich in [WHUP04] für eine zweidimensionale Positionsbestimmung und in [WHU04] lediglich zur Richtungsbestimmung eingesetzt. Dabei konnte gezeigt werden, dass eine genauere Sprecherverfolgung mittels der Kombination aus PCA *Beamforming* und Partikel-Filterung im Vergleich zur Kombination aus GCC bzw. DSB-SRP und Partikel-Filterung erreicht wird. Außerdem wurde in [WHUP04, WHU04] die Überlegenheit der Partikel-Filterung gegenüber dem Kalman Filter für diese Anwendung demonstriert.

7.4 Zusammenfassung

In diesem Kapitel wurden zwei häufig benutzte Verfahren zur Sprecherrichtungsbestimmung vorgestellt. Dies ist zum einen die TDOA-Methode, welche die Zeitdifferenz zweier Mikrophonsignale bestimmt und zum anderen das SRP-Verfahren, welches die Ausgangsleistung eines *Beamformers* durch Steuerung seiner Richtcharakteristik maximiert.

Diese Methoden wurden hier derart erweitert, dass auch unter Einfluss starker stationärer Störungen eine gute Sprecherrichtungsbestimmung möglich ist. Dabei erfolgt nicht die Auswertung der Mikrophonsignale, sondern der iterativ bestimmten dominanten Eigenvektoren des verallgemeinerten Eigenwertproblems. Bei einer guten Schätzung der spektralen Kreuzleistungsdichten der Störung kann somit auch eine zuverlässige Richtungsschätzung in Anwesenheit von gerichteten Störquellen erfolgen.

Da die Abtastung der Richtcharakteristik der Eigenvektoren sehr rechenintensiv ist, sind Implementierungsmöglichkeiten aufgezeigt worden, die zu einer erheblichen Reduzierung des Berechnungsaufwands führen. Die Komplexität ist dadurch ähnlich wie die der Korrelationsmethode, bei einer nur sehr geringen damit einhergehenden Verzögerung, und dennoch einem insgesamt robusteren Verhalten im Vergleich zur Korrelationsmethode.

Kapitel 8

GEV-Beamformer in GSC-Struktur

In Kapitel 4 wurde das Prinzip des statistisch optimalen *Beamformings* aufgezeigt, wobei das Optimierungskriterium in Abschnitt 4.2 aus der Minimierung der Varianz des Ausgangssignals des *Beamformers* unter der Einhaltung einer Nebenbedingung besteht. Basierend auf diesem Ansatz ist in dem bekannten Verfahren nach [Fro72] zur Minimierung der Rauschleistung ein Adaptionsschema mit Nebenbedingung beschrieben. In [GJ82] wurde das Minimierungsproblem mit Nebenbedingung umgewandelt in ein Minimierungsproblem ohne Nebenbedingung, so dass die adaptiven Filter zur Störgeräuschreduktion (engl. *Adaptive Noise Cancellation*, ANC) einfach mittels LMS-Algorithmus realisiert werden können. Dabei erfolgt die Minimierung des Rauschens in einem Signal, welches mittels eines unveränderlichen *Beamformers* (engl. *Fixed Beamformer*, FBF) erzeugt wird. Hierbei wird davon ausgegangen, dass das mehrkanalige Eingangssignal des FBFs bezüglich des Sprachsignals zeitangepasst, also kohärent vorliegt. Die adaptiven Filter benötigen dann am Eingang vorverarbeitete Mikrophonsignale, die möglichst keine Sprachkomponenten mehr enthalten und daher auch als Störgeräuschreferenzsignale bezeichnet werden. Die Störgeräuschreferenzsignale gehen prinzipiell aus einer Matrixmultiplikation mit den Mikrophonsignalen hervor, wobei diese Sprachsignal-blockierende Matrix (engl. *Blocking Matrix*, BM) nach [GJ82] eine feste, nicht adaptive[1] Struktur aufweist. Die sich ergebende Gesamtstruktur bestehend aus FBF, BM und ANC wird als *Generalized Sidelobe Canceller* (GSC) bezeichnet, siehe Bild 8.1.

Die Leistungsfähigkeit eines GSCs zur Störgeräuschreduktion hängt insbesondere von der Güte der Störgeräuschreferenzsignale ab, welche möglichst frei von dem Nutzsignal sein sollten. Diese Eigenschaft wird dabei maßgeblich durch zwei Problemstellungen beeinflusst: zum einen ist dies die Mehrwegeausbreitung des Sprachsignals und zum anderen simultan aktive Störgeräuschquellen.

In diesem Kapitel werden unterschiedliche Realisierungen der *Blocking Matrix* behandelt. Dabei wird insbesondere ein neuartiges Verfahren vorgestellt, welches auf einem GEV *Beamforming* basiert. Dieses hat den Vorteil, sich adaptiv dem Sprachsignal anzupassen, auch wenn ein permanentes Störschallfeld vorliegt.

[1] Ein feste, nicht adaptive *Blocking Matrix* setzt eine der Sprecherposition entsprechende Laufzeitkompensation des direkten Pfades voraus.

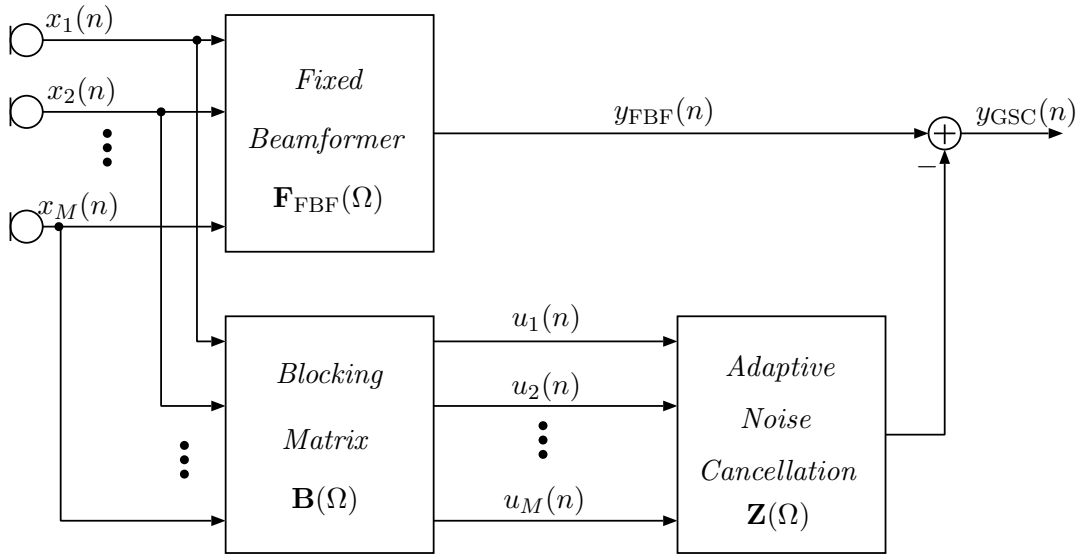

Bild 8.1: Blockschaltbild des *Generalized Sidelobe Cancellers*.

8.1 GSC in stationärer Umgebung

Eine äquivalente Schreibweise zur Minimierung der Kostenfunktion Gl. (4.21) ist gegeben durch

$$\underset{\mathbf{F}(\Omega)}{\text{minimiere}} \qquad \mathbf{F}^H(\Omega)\mathbf{\Phi_{XX}}(\Omega)\mathbf{F}(\Omega) \tag{8.1}$$

$$\text{mit} \qquad \mathbf{F}^H(\Omega)\mathbf{H}(\Omega) = W(\Omega), \tag{8.2}$$

mit der spektralen Gewichtung $W(\Omega)$ (vgl. Gl. (4.20)). Dieser Ansatz kann mit Hilfe der Lagrange-Funktion und einem Gradienten-Abstiegs-Verfahrens gelöst werden (siehe Lösung Gl. (4.28)). Für eine unverzerrte Filterung des Sprachsignals muss folgende Bedingung gelten

$$W(\Omega) = 1. \tag{8.3}$$

Optimale Filter der ANC

Nun wird der Filterkoeffizientenvektor aufgespaltet in zwei additive Anteile

$$\mathbf{F}(\Omega) = \mathbf{F}_{\mathrm{FBF}}(\Omega) - \mathbf{B}(\Omega)\mathbf{Z}(\Omega), \tag{8.4}$$

wobei $\mathbf{F}_{\mathrm{FBF}}(\Omega)$ die eigentliche Strahlformung (*Fixed Beamformer*), $\mathbf{B}(\Omega)$ die Sprachsignal-Blockierung (*Blocking Matrix*) und $\mathbf{Z}(\Omega)$ die Störgeräusch-Auslöschung (*Noise Cancellation*) beschreibt (vgl. Bild 8.1). Nach Einsetzen von Gl. (8.4) in Gl. (8.2) mit $W(\Omega) = 1$ ergibt sich

$$\left[\mathbf{F}_{\mathrm{FBF}}^H(\Omega) - \mathbf{Z}^H(\Omega)\mathbf{B}^H(\Omega)\right]\mathbf{H}(\Omega) = 1, \tag{8.5}$$

wobei Gl. (8.5) durch die Bedingungen

$$\mathbf{F}_{\mathrm{FBF}}^H(\Omega)\mathbf{H}(\Omega) = 1 \tag{8.6}$$

$$\mathbf{B}^H(\Omega)\mathbf{H}(\Omega) = 0 \tag{8.7}$$

erfüllt werden kann. Falls Gl. (8.6) und Gl. (8.7) eingehalten werden, können die mehrkanaligen Filter $\mathbf{Z}(\Omega)$ zur Erfüllung der Bedingung Gl. (8.5) beliebig gewählt werden und müssen keine Nebenbedingung einhalten. Daher sind sie nun so zu wählen, dass in dem einkanaligen Ausgangssignal des *Fixed Beamformers*

$$Y_{\mathrm{FBF}}(\Omega) = \mathbf{F}_{\mathrm{FBF}}^{H}(\Omega)\mathbf{X}(\Omega), \tag{8.8}$$

alle Störsignalkomponenten, welche mit dem mehrkanaligen Störgeräuschreferenzsignal

$$\mathbf{U}(\Omega) = \mathbf{B}^{H}(\Omega)\mathbf{X}(\Omega) \tag{8.9}$$

räumlich korreliert sind, entfernt werden, und sich das letztendliche Ausgangssignal des GSCs zu

$$Y_{\mathrm{GSC}}(\Omega) = Y_{\mathrm{FBF}}(\Omega) - \mathbf{Z}^{H}(\Omega)\mathbf{U}(\Omega) \tag{8.10}$$

ergibt. Die Kostenfunktion für das Minimierungsproblem ist

$$J_{\mathrm{GSC}}(\mathbf{Z}(\Omega)) = \left[\mathbf{F}_{\mathrm{FBF}}^{H}(\Omega) - \mathbf{Z}^{H}(\Omega)\mathbf{B}^{H}(\Omega)\right] \mathbf{\Phi}_{\mathbf{XX}}(\Omega) \left[\mathbf{F}_{\mathrm{FBF}}(\Omega) - \mathbf{B}(\Omega)\mathbf{Z}(\Omega)\right] \tag{8.11}$$

und ergibt somit den Gradientenvektor

$$\nabla_{\mathbf{Z}} J_{\mathrm{GSC}}(\mathbf{Z}(\Omega)) = -\mathbf{B}^{H}(\Omega)\mathbf{\Phi}_{\mathbf{XX}}(\Omega)\mathbf{F}_{\mathrm{FBF}}(\Omega) + \mathbf{B}^{H}(\Omega)\mathbf{\Phi}_{\mathbf{XX}}(\Omega)\mathbf{B}(\Omega)\mathbf{Z}(\Omega). \tag{8.12}$$

Durch Nullsetzen von Gl. (8.12) kann das mehrkanalige Wiener Filter mit den optimalen Koeffizienten angegeben werden als

$$\begin{aligned} \mathbf{Z}_{\mathrm{opt}}(\Omega) &= \left[\mathbf{B}^{H}(\Omega)\mathbf{\Phi}_{\mathbf{XX}}(\Omega)\mathbf{B}(\Omega)\right]^{-1} \mathbf{B}^{H}(\Omega)\mathbf{\Phi}_{\mathbf{XX}}(\Omega)\mathbf{F}_{\mathrm{FBF}}(\Omega) && (8.13)\\ &= \mathbf{\Phi}_{\mathbf{UU}}^{-1}(\Omega)\mathbf{\Phi}_{\mathbf{U}Y_{\mathrm{FBF}}}(\Omega), && (8.14) \end{aligned}$$

wobei $\mathbf{\Phi}_{\mathbf{UU}}(\Omega) = E\{\mathbf{U}(\Omega)\mathbf{U}^{H}(\Omega)\}$ als invertierbar angenommen wird und $\mathbf{\Phi}_{\mathbf{U}Y_{\mathrm{FBF}}}(\Omega) = E\{\mathbf{U}(\Omega)Y_{\mathrm{FBF}}^{*}(\Omega)\}$ gilt.

Implementierung und Adaption der ANC

Die Filterkoeffizienten Gl. (8.13) der mehrkanaligen ANC können in einer Implementierung iterativ für jeden Verarbeitungsblock mit dem Index m und jede Frequenzkomponente Ω_k über die normalisierte LMS-Adaptionsregel bestimmt werden

$$\begin{aligned} \mathbf{Z}_{m+1}(\Omega_k) &= \mathbf{Z}_{m}(\Omega_k) + \mu P_{m}^{-1}(\Omega_k)\mathbf{U}_{m}(\Omega_k)Y_{\mathrm{GSC}\,m}^{*}(\Omega_k) && (8.15)\\ P_{m+1}(\Omega_k) &= \alpha P_{m}(\Omega_k) + (1-\alpha)M^{-1}\mathbf{U}_{m}^{H}(\Omega_k)\mathbf{U}_{m}(\Omega_k), && (8.16) \end{aligned}$$

mit der festen Schrittweite μ und dem Glättungsfaktor α. Im Sinne der adaptiven Filterung beschreibt $Y_{\mathrm{GSC}\,m}(\Omega_k)$ das Fehlersignal zwischen dem Referenzsignal des FBFs und dem gefilterten Signal am Ausgang der ANC. Dieses kann genau genommen nur für Signalblöcke herangezogen werden, in denen kein Nutzsignal enthalten ist. Daher sollte die Adaptionsregel Gl. (8.15) und Gl. (8.16) über eine Sprache/Pause-Detektion gesteuert werden. Außerdem ist der Fehler im Zeitbereich zu ermitteln, und durch Einfügen von Nullen in die Filterimpulsantworten werden zyklische Effekte vermieden [Shy92]. Um eine möglichst hohe Störgeräuschreduktion zu erhalten, sollte jeweils die Filterlänge möglichst groß gewählt werden. Dies bedeutet jedoch auch, dass die Adaptionsdauer zunimmt. Bei einer Abtastrate von $f_{Ab} = 12\,\mathrm{kHz}$ stellt eine Filterlänge von 1024 einen guten Kompromiss und sinnvollen Wert dar.

Allgemeine Form der BM

Das Ziel der *Blocking Matrix* ist, eine Projektion der Eingangssignale auf den zur Sprachsignalkomponente orthogonalen Unterraum durchzuführen. Für die Einhaltung der Bedingung $\mathbf{B}^H(\Omega)\mathbf{H}(\Omega) = 0$ wird eine Struktur in der Form

$$\mathbf{B}^H(\Omega) = \mathbf{I} - \boldsymbol{\mathcal{B}}^H(\Omega) \tag{8.17}$$

gewählt, wobei $\boldsymbol{\mathcal{B}}^H(\Omega)\mathbf{H}(\Omega) = \mathbf{H}(\Omega)$ gelten soll. Die Projektion $\boldsymbol{\mathcal{B}}^H(\Omega)$ soll das Sprachsignal also so gut wie möglich rekonstruieren. Es kann folglich die allgemeine Formulierung

$$\boldsymbol{\mathcal{B}}^H(\Omega) = \frac{\mathbf{H}(\Omega)\boldsymbol{\mathcal{W}}^H(\Omega)}{\boldsymbol{\mathcal{W}}^H(\Omega)\mathbf{H}(\Omega)} \tag{8.18}$$

verwendet werden, wobei der Vektor $\boldsymbol{\mathcal{W}}(\Omega)$ in Gl. (8.18) prinzipiell beliebig gewählt werden kann aber ungleich dem Nullvektor sein muss und nicht orthogonal zu $\mathbf{H}(\Omega)$ sein darf. Es ist also direkt zu sehen, dass mit der Formulierung Gl. (8.18) die Bedingung Gl. (8.7) eingehalten wird. Für die Filterung des Eingangssignals $\mathbf{X}(\Omega) = S_c(\Omega)\mathbf{H}(\Omega) + \mathbf{N}(\Omega)$ mit der *Blocking Matrix* ergibt sich

$$\mathbf{U}(\Omega) = \mathbf{B}^H(\Omega)\mathbf{X}(\Omega) = \left[\mathbf{I} - \frac{\mathbf{H}(\Omega)\boldsymbol{\mathcal{W}}^H(\Omega)}{\boldsymbol{\mathcal{W}}^H(\Omega)\mathbf{H}(\Omega)}\right]\mathbf{N}(\Omega), \tag{8.19}$$

wobei offensichtlich das Sprachsignal verschwindet und in $\mathbf{U}(\Omega)$ nur noch gefilterte Störsignalkomponenten verbleiben.

Es stellt sich nun die Frage, wie die Matrix $\boldsymbol{\mathcal{B}}^H(\Omega)$ realisiert werden soll. Wie ist also der Vektor $\boldsymbol{\mathcal{W}}(\Omega)$ zu wählen und wie kann die Raumübertragungsfunktion $\mathbf{H}(\Omega)$ bestimmt werden.

8.2 Realisierung der Blocking Matrix

Im Folgenden sollen drei BM-Varianten aus der Literatur vorgestellt werden:

- die einfache Methode der Subtraktion zeitangepasster Mikrophonsignale nach Griffiths und Jim [GJ82],
- die Lösung nach Gannot et al. [GBW01] durch Einsetzen von zuvor bestimmten Verhältnissen der Raumübertragungsfunktionen und
- das robuste Verfahren nach Hoshuyama et al. [HSH99], bei dem die Sprachanteile in den Mikrophonsignalen mittels adaptiver Filter und einem Sprachreferenzsignal entfernt werden.

Weiterhin wird eine neuartige Realisierung basierend auf dem GEV *Beamforming* hergeleitet.

Zunächst sollen zwei fundamentale Realisierungen der *Blocking Matrix* aufgezeigt werden. Wählt man $\boldsymbol{\mathcal{W}}^H(\Omega) = (1, 0, \ldots, 0)$ so ergibt sich

$$\mathbf{B}^H_{\mathrm{TFR}}(\Omega) = \frac{1}{H_1(\Omega)} \begin{bmatrix} 0 & 0 & \ldots & & 0 \\ -H_2(\Omega) & H_1(\Omega) & 0 & \ldots & \vdots \\ -H_3(\Omega) & 0 & H_1(\Omega) & \ldots & \\ \vdots & \vdots & & \ddots & \\ -H_M(\Omega) & 0 & \ldots & & H_1(\Omega) \end{bmatrix}. \tag{8.20}$$

Aufgrund der Tatsache, dass in Gl. (8.20) die Verhältnisse $H_i(\Omega)/H_1(\Omega)$ für $i = 2, 3, \ldots, M$ zu bestimmen sind, wird die Matrix auch *Transfer Function Ratio* (TFR) *Blocking Matrix* (TFRBM) genannt und führt daher zu dem Index "TFR" in Gl. (8.20). Das Grundprinzip bei der Filterung der Mikrophonsignale mit $\mathbf{B}^H_{\mathrm{TFR}}(\Omega)$ besteht darin, paarweise aufeinander angepasste Signale zu subtrahieren, also $X_i(\Omega) - H_i(\Omega)/H_1(\Omega)X_1(\Omega)$ für $i = 2, 3, \ldots, M$ zu berechnen.

Als nächstes ergibt sich mit $\boldsymbol{\mathcal{W}}^H(\Omega) = (1, 1, \ldots, 1)$ die voll besetzte Matrix

$$\mathbf{B}^H_{\mathrm{TF}}(\Omega) = \frac{1}{\sum_{i=1}^{M} H_i(\Omega)} \begin{bmatrix} \sum_{i=2}^{M} H_i(\Omega) & -H_1(\Omega) & \ldots & -H_1(\Omega) \\ -H_2(\Omega) & \sum_{i=1,\, i\neq 2}^{M} H_i(\Omega) & \ldots & -H_2(\Omega) \\ \vdots & & \ddots & \\ -H_M(\Omega) & -H_M(\Omega) & \ldots & \sum_{i=1}^{M-1} H_i(\Omega) \end{bmatrix} \tag{8.21}$$

mit der Bezeichnung "TF" für *Transfer Function*. Die Matrix in Gl. (8.21) soll demzufolge *Transfer Function Blocking Matrix* (TFBM) genannt werden. Der Rang von $\mathbf{B}^H_{\mathrm{TF}}(\Omega)$ ist weiterhin $M - 1$, was bedeutet, dass eins der M Störgeräuschreferenzsignale linear abhängig ist von den anderen Störgeräuschreferenzsignalen.

8.2.1 BM nach Griffiths und Jim

Die Grundidee nach [GJ82] basiert auf der Annahme der Freifeldausbreitung des Sprachsignals, so dass lediglich zeitangepasste Mikrophonsignale subtrahiert werden müssen, um das Nutzsignal zu entfernen. Die Übertragungsfunktion für die Sprecherrichtung θ_s soll also beschrieben sein durch

$$\mathbf{d}(\Omega, \theta_s) = (e^{j\Omega\tau_1(\theta_s)/T},\, e^{j\Omega\tau_2(\theta_s)/T}, \ldots, e^{j\Omega\tau_M(\theta_s)/T})^H. \tag{8.22}$$

Bei einer Implementierung würden die durch Gl. (8.22) entstehenden relativen Verzögerungen in einem ersten Schritt kompensiert werden

$$\tilde{\mathbf{X}}(\Omega) = e^{-j\Omega t_k/T}\, \mathrm{diag}\{(\mathrm{e}^{\mathrm{j}\Omega\tau_1(\theta_s)/\mathrm{T}},\, \mathrm{e}^{\mathrm{j}\Omega\tau_2(\theta_s)/\mathrm{T}}, \ldots, \mathrm{e}^{\mathrm{j}\Omega\tau_\mathrm{M}(\theta_s)/\mathrm{T}})\}\, \mathbf{X}(\Omega), \tag{8.23}$$

wobei die Verzögerung $e^{-j\Omega t_k/T}$ mit $t_k > \max\{\tau_i\}$ zur Realisierung einer kausalen Filterung eingefügt wurde. Das so kohärent verschobene mehrkanalige Signal $\tilde{\mathbf{X}}(\Omega)$ dient als Eingangssignal für die *Blocking Matrix*.

Mit diesen Annahmen ergibt sich aus Gl. (8.20) die einfache Form der *Delay Only Ratio Blocking Matrix* (DORBM) mit dem Index "DOR"

$$\mathbf{B}^H_{\mathrm{DOR}}(\Omega) = \begin{bmatrix} 0 & 0 & \ldots & & 0 \\ -1 & 1 & 0 & \ldots & 0 \\ -1 & 0 & 1 & & \\ \vdots & \vdots & & \ddots & \\ -1 & 0 & 0 & \ldots & 1 \end{bmatrix}, \tag{8.24}$$

und entsprechend aus Gl. (8.21) folgt die *Delay Only Blocking Matrix* (DOBM) mit dem Index "DO"

$$\mathbf{B}_{\mathrm{DO}}^{H}(\Omega) = \frac{1}{M}\begin{bmatrix} M-1 & -1 & \dots & -1 \\ -1 & M-1 & & \vdots \\ \vdots & & \ddots & \\ -1 & -1 & \dots & M-1 \end{bmatrix}. \tag{8.25}$$

In Gl. (8.25) wird also quasi von jedem Eingangssignal der Mittelwert der anderen Eingangssignale subtrahiert. Auch hier ist der Rang der $(M \times M)$-Matrix $\mathbf{B}_{\mathrm{DO}}^{H}(\Omega)$ wieder $M-1$.

Implementierung der DORBM und DOBM

Grundsätzlich sind die Matrizen Gl. (8.24) und Gl. (8.25) nichtadaptiv und benötigen also kein direktes Nachführen von Koeffizienten. Die Subtraktion kann sehr effizient und ohne Verzerrungen im Zeitbereich umgesetzt werden. Bei der Implementierung des Gesamtsystems ist jedoch eine adaptive Sprecherrichtungsbestimmung und -nachführung, sowie eine Laufzeitkompensation des direkten Ausbreitungspfades notwendig. Wird der Zeitausgleich korrekt vorgenommen, besteht aufgrund der nichtadaptiven Struktur der BM der Vorteil einer störgeräuschunabhängigen Sprachsignalunterdrückung. Da aber die Annahme der Freifeldausbreitung für reale Anwendungen in verhallten Räumen nicht haltbar ist, gelangt mit steigender Nachhallzeiten ein wachsender Anteil an Sprachsignalkomponenten in die Störgeräuschreferenzsignale hinein. Weiterhin entsteht dieser Effekt natürlich auch bei nicht korrekt zeitangepassten Mikrophonsignalen.

8.2.2 BM nach Gannot et al.

Für die TFR *Blocking Matrix* Gl. (8.20) müssen die Verhältnisse $H_i(\Omega)/H_1(\Omega)$ für $i = 2, 3, \dots, M$ geschätzt werden. Ein Verfahren hierzu ist in [GBW99] erstmals im Zusammenhang mit einer GSC-Realisierung vorgestellt worden, wobei ausführlichere Beschreibungen in [Gan00, GBW01] zu finden sind. Grundlage bildet Gl. (8.9), welche umgestellt wird zu

$$X_i(\Omega) = U_{i-1}(\Omega) + \frac{H_i(\Omega)}{H_1(\Omega)} X_1(\Omega). \tag{8.26}$$

Mit Gl. (8.26) wird unter Beachtung der blockweisen Verarbeitung eine gleichgewichtete Schätzung der spektralen Kreuzleistungsdichte zwischen dem i-ten und dem ersten Mikrophon für den Zeitpunkt m angegeben zu

$$\hat{\phi}_{X_i X_1,m}^{(\mathrm{GG})}(\Omega) = \hat{\phi}_{U_{i-1} X_1,m}^{(\mathrm{GG})}(\Omega) + \frac{H_i(\Omega)}{H_1(\Omega)} \hat{\phi}_{X_1 X_1,m}^{(\mathrm{GG})}(\Omega), \quad i = 2, 3, \dots, M, \tag{8.27}$$

wobei ausgenutzt wurde, dass das Nutz- und das Störsignal miteinander unkorreliert und jeweils mittelwertfrei sind.

Weiterhin wird der Fehler zwischen dem Spektrum des $(i-1)$-ten Ausgangssignal der BM und dem ersten Mikrophonsignal definiert

$$\mathcal{E}_{i-1,m}(\Omega) = \hat{\phi}_{U_{i-1} X_1,m}^{(\mathrm{GG})}(\Omega) - \phi_{U_{i-1} X_1}(\Omega). \tag{8.28}$$

Mit Gl. (8.27) und Gl. (8.28) ist es möglich, nach N Blöcken folgendes überbestimmtes lineares Gleichungssystem aufzustellen

$$\begin{bmatrix} \hat{\phi}^{(\mathrm{GG})}_{X_iX_1,1}(\Omega) \\ \hat{\phi}^{(\mathrm{GG})}_{X_iX_1,2}(\Omega) \\ \vdots \\ \hat{\phi}^{(\mathrm{GG})}_{X_iX_1,N}(\Omega) \end{bmatrix} = \begin{bmatrix} \hat{\phi}^{(\mathrm{GG})}_{X_1X_1,1}(\Omega) & 1 \\ \hat{\phi}^{(\mathrm{GG})}_{X_1X_1,2}(\Omega) & 1 \\ \vdots & \\ \hat{\phi}^{(\mathrm{GG})}_{X_1X_1,N}(\Omega) & 1 \end{bmatrix} \begin{bmatrix} H_i(\Omega)/H_1(\Omega) \\ \phi_{U_{i-1}X_1}(\Omega) \end{bmatrix} + \begin{bmatrix} \mathcal{E}_{i-1,1}(\Omega) \\ \mathcal{E}_{i-1,2}(\Omega) \\ \vdots \\ \mathcal{E}_{i-1,N}(\Omega) \end{bmatrix}. \tag{8.29}$$

Mit der entscheidenden Forderung der Stationarität des Störsignals und der Ausnutzung der Nichtstationarität der Sprache kann eine Schätzung $\hat{H}_i(\Omega)/\hat{H}_1(\Omega)$ abgeleitet werden. Dabei wird die Methode der kleinsten Quadrate auf das Gleichungssystem Gl. (8.29) nach dem in [SW96] vorgestellten Prinzip angewendet. Die Lösung ergibt sich dann laut [GBW01] zu

$$\frac{\hat{H}_i(\Omega)}{\hat{H}_1(\Omega)} = \frac{\sum\limits_{m=1}^{N} \left(\hat{\phi}^{(\mathrm{GG})}_{X_1X_1,m}(\Omega) \hat{\phi}^{(\mathrm{GG})}_{X_iX_1,m}(\Omega) \right) - \sum\limits_{m=1}^{N} \hat{\phi}^{(\mathrm{GG})}_{X_1X_1,m}(\Omega) \sum\limits_{m=1}^{N} \hat{\phi}^{(\mathrm{GG})}_{X_iX_1,m}(\Omega)}{\sum\limits_{m=1}^{N} \left(\hat{\phi}^{(\mathrm{GG})}_{X_1X_1,m}(\Omega) \right)^2 - \sum\limits_{m=1}^{N} \left(\hat{\phi}^{(\mathrm{GG})}_{X_iX_1,m}(\Omega) \right)^2}. \tag{8.30}$$

Implementierung der TFRs

Grundsätzlich ist die Implementierung von Gl. (8.30) für diskrete Spektralkomponenten Ω_k vorzunehmen; es werden also die Verhältnisse $\hat{H}_i(\Omega_k)/\hat{H}_1(\Omega_k)$ bestimmt. In [GBW01] wird berichtet, dass die Blöcke zur gleichgewichteten Schätzung der Kreuzleistungsdichten sich nicht überlappen sollten. Weiterhin ist natürlich die Schätzung durchzuführen, wenn das Nutzsignal auch in den Mikrophonsignalen vorliegt, weshalb eine Sprache/Pause-Detektion notwendig ist. In der Realisierung [GBW01] wurden die Filterlängen in der *Blocking Matrix* jeweils zu 181 bei einer Abtastrate von 8 kHz gewählt. Daher scheint eine Wahl von $B = 256$ für die Filterimpulsantworten bei der Abtastrate $f_{Ab} = 12\,\mathrm{kHz}$ gerechtfertigt zu sein. Diese sind wie folgt zu ermitteln. Aus den nichtüberlappenden Abtastwerten am Eingang werden $L = 512$ Daten mittels einer Hamming-Fensterung herausgenommen und im Frequenzbereich entsprechend viele Koeffizienten mittels Gl. (8.30) berechnet. Nach der Rücktransformation in den Zeitbereich werden $B = 256$ Koeffizienten herausgeschnitten[2], mit Nullen auf eine Länge $L = 512$ aufgefüllt und wieder in den Frequenzbereich transformiert.

Für eine konsequente Nutzung der Verhältnisse der Übertragungsfunktionen können diese auch in den FBF eingesetzt werden. Die entstehenden Sprachverzerrungen des Gesamtsystems sind ausführlich in [GBW04] behandelt. Dabei scheinen insbesondere in dem unteren Frequenzbereich ($f < 500\,\mathrm{Hz}$) Probleme aufzutreten.

Die GSC-Struktur kann zur weiteren Störgeräuschreduktion mit einem *Post Filter* [GC04] und einer Echokompensation [RGC07a] erweitert werden. In [RGC07b] ist das Gesamtsystem schließlich noch auf das Vorhandensein eines zusätzlichen Sprechers ausgelegt worden, also einem Szenario mit zwei instationären Quellen.

[2] Prinzipiell lässt sich zur Vermeidung zyklischer Effekte bei der Filterung im Frequenzbereich auch folgende Methode verwenden: Nach der Fourier-Transformation der Länge 512 werden zu jeder zweiten Frequenzkomponente die Verhältnisse Gl. (8.30) berechnet. Diese 256 Koeffizienten sind in den Zeitbereich zu transformieren und mit Nullen zu erweitern, so dass schließlich wieder eine Fourier-Transformation der Länge 512 angewendet werden kann.

8.2.3 BM nach Hoshuyama et al.

Im Folgenden wird eine Variante der BM beschrieben, die ohne direkte Berechnung der Übertragungsfunktionen bzw. Verhältnisse dieser auskommt. Das Verfahren wurde erstmals in [HSH96] vorgestellt, wobei die Sprachanteile in den Mikrophonsignalen mittels adaptiver Filter und einem Sprachreferenzsignal entfernt werden. Eine genauere Beschreibung ist in [HSH99] zu finden. Die dort vorgestellte LMS-Adaption ist im Zeitbereich realisiert und in [HK01] auf eine recheneffiziente Version im Frequenzbereich übertragen worden.

Die Idee besteht darin, ein Sprachreferenzsignal $Y_{\text{ref}}(\Omega)$ zu erzeugen, welches aus der Filterung der Eingangsdaten mit dem Filtervektor $\mathbf{F}_{\text{ref}}(\Omega)$ hervorgeht

$$Y_{\text{ref}}(\Omega) = \mathbf{F}_{\text{ref}}^{H}(\Omega)\mathbf{X}(\Omega). \tag{8.31}$$

Zwischen diesem Referenzsignal und den Eingangssignalen werden weitere FIR-Filter $\boldsymbol{\mathcal{G}}(\Omega)$ eingefügt, um die Störgeräuschreferenzsignale zu generieren

$$\mathbf{U}(\Omega) = \mathbf{X}(\Omega) - \boldsymbol{\mathcal{G}}(\Omega)Y_{\text{ref}}(\Omega). \tag{8.32}$$

Die statistisch optimalen Koeffizienten sollen mit optimalen Eingangsdaten, also $\mathbf{X}(\Omega) = \mathbf{S}(\Omega)$ und dem optimalen Referenzsignal

$$Y_{\text{opt}}(\Omega) = Y_{\text{ref}}(\Omega)\Big|_{\mathbf{X}(\Omega)=\mathbf{S}(\Omega)} \tag{8.33}$$

bestimmt werden, mittels der Bedingung

$$E\left\{\big(\mathbf{X}(\Omega) - \boldsymbol{\mathcal{G}}(\Omega)Y_{\text{ref}}(\Omega)\big)Y_{\text{ref}}^{*}(\Omega)\right\}\Big|_{\mathbf{X}(\Omega)=\mathbf{S}(\Omega)} \stackrel{!}{=} 0. \tag{8.34}$$

Das Ergebnis ist das folgende Wiener Filter

$$\boldsymbol{\mathcal{G}}_{\text{opt}}(\Omega) = \frac{\boldsymbol{\Phi}_{\mathbf{S}Y_{\text{opt}}}(\Omega)}{\phi_{Y_{\text{opt}}Y_{\text{opt}}}(\Omega)}. \tag{8.35}$$

Die optimalen Filterkoeffizienten des Wiener Filters Gl. (8.35) können weiter umgeformt werden zu

$$\begin{aligned}\boldsymbol{\mathcal{G}}_{\text{opt}}(\Omega) &= \frac{\phi_{S_cS_c}(\Omega)\mathbf{H}(\Omega)\mathbf{H}^{H}(\Omega)\mathbf{F}_{\text{ref}}(\Omega)}{\phi_{S_cS_c}(\Omega)\mathbf{H}^{H}(\Omega)\mathbf{F}_{\text{ref}}(\Omega)\mathbf{F}_{\text{ref}}^{H}(\Omega)\mathbf{H}(\Omega)} &(8.36)\\ &= \frac{\mathbf{H}(\Omega)}{\mathbf{F}_{\text{ref}}^{H}(\Omega)\mathbf{H}(\Omega)}. &(8.37)\end{aligned}$$

An Gl. (8.37) kann abgelesen werden, dass mit Hilfe der idealisierten Annahme $\mathbf{X}(\Omega) = \mathbf{S}(\Omega)$ gerade eine Systemidentifikation möglich ist, da eine skalierte Version der Raumübertragungsfunktion bestimmt wurde.

Da durch die Subtraktion in Gl. (8.32) die Störgeräuschreferenzsignale mittels einer adaptiven Sprachsignalauslöschung (engl. *Adaptive Speech Cancellation*, ASC) generiert werden sollen, wird die so entstehende *Blocking Matrix* in dieser Arbeit als ASCBM bezeichnet und mit dem Index "ASC" versehen. Die optimale ASCBM ergibt sich aus den oberen Erkenntnissen zu

$$\begin{aligned}\mathbf{B}_{\text{ASC opt}}^{H}(\Omega) &= \mathbf{I} - \boldsymbol{\mathcal{G}}_{\text{opt}}(\Omega)\mathbf{F}_{\text{ref}}^{H}(\Omega) &(8.38)\\ &= \mathbf{I} - \frac{\mathbf{H}(\Omega)\mathbf{F}_{\text{ref}}^{H}(\Omega)}{\mathbf{F}_{\text{ref}}^{H}(\Omega)\mathbf{H}(\Omega)}. &(8.39)\end{aligned}$$

Bei dem Vergleich von Gl. (8.39) mit der allgemeinen Formulierung Gl. (8.17) und Gl. (8.18) gilt für diesen Ansatz offensichtlich

$$\boldsymbol{\mathcal{W}}(\Omega) = \mathbf{F}_{\mathrm{ref}}(\Omega). \tag{8.40}$$

Grundsätzlich gilt auch hier wieder, dass die Wahl des Vektors $\mathbf{F}_{\mathrm{ref}}(\Omega)$ beliebig ist, solange dieser ungleich dem Nullvektor und nicht orthogonal zu der Raumübertragungsfunktion des Sprachsignals ist. Geht man zunächst noch davon aus, dass das Eingangssignal keine Störkomponenten beinhaltet, so führt z. B. die Wahl von $\mathbf{F}_{\mathrm{ref}} = (1, 0, \ldots, 0)^T$ zu einer BM die identisch zu $\mathbf{B}_{\mathrm{TFR}}(\Omega)$ aus Gl. (8.20) ist. Bei der realen Anwendung gilt jedoch $\mathbf{X}(\Omega) = \mathbf{S}(\Omega) + \mathbf{N}(\Omega)$, weshalb eine andere Wahl für $\mathbf{F}_{\mathrm{ref}}(\Omega)$ zur Erzeugung eines Sprachreferenzsignals sinnvoll ist.

Da der Fokus der Arbeiten von z. B. [HSH99, HS01] und [HK03] auf der Unterdrückung nicht stationärer Quellen – also weiterer Sprecher – liegt, wird nur von unkorreliertem Mikrophonrauschen und sehr geringem diffusen Rauschen ausgegangen. Diese Rauschkomponenten können mit dem FBF in der Realisierung als DSB oder besser z. B. mit einem Dolph-Chebyshev-Fenster angewendet auf die zeitkompensierten Mikrophonsignale deutlich reduziert werden. Daher kann das Ausgangssignal $Y_{\mathrm{FBF}}(\Omega)$ des FBFs als Sprachreferenzsignal dienen und für die Filterkoeffizienten gilt demnach

$$\mathbf{F}_{\mathrm{ref}}(\Omega) = \mathbf{F}_{\mathrm{FBF}}(\Omega). \tag{8.41}$$

Implementierung und Adaption der ASCBM

Die blockorientierte adaptive Bestimmung von $\boldsymbol{\mathcal{G}}_m(\Omega_k)$ für die diskreten Spektralkomponenten Ω_k und den Verarbeitungsblock m ergibt sich demzufolge äquivalent zu Gl. (8.15) durch

$$\begin{aligned} \boldsymbol{\mathcal{G}}_{m+1}(\Omega_k) &= \boldsymbol{\mathcal{G}}_m(\Omega_k) + \mu P_m^{-1}(\Omega_k)\mathbf{X}_m(\Omega_k)Y^*_{\mathrm{FBF}\ m}(\Omega_k) && (8.42)\\ P_{m+1}(\Omega_k) &= \alpha P_m(\Omega_k) + (1-\alpha)|Y_{\mathrm{FBF}\ m}(\Omega_k)|^2, && (8.43) \end{aligned}$$

wiederum mit der festen Schrittweite μ und dem Glättungsfaktor α. Auch hier ist wieder auf die Besonderheiten der Filterung im Frequenzbereich zu achten [Shy92]. Robustheitsaspekte wie eine Begrenzung der Filterkoeffizienten oder ein *Leaky*-Faktor sind in Gl. (8.42) nicht berücksichtigt worden. Im Gegensatz zu Gl. (8.15) erfolgt die Iteration in Gl. (8.42) während Sprachaktivität, welche über eine Sprache/Pause-Detektion angezeigt werden muss. Nach der Analyse [HK02] zu der ASC *Blocking Matrix* ist bei einer Abtastrate von $f_{Ab} = 12\,\mathrm{kHz}$ eine Länge von 256 für die adaptiven Filter als sinnvoll zu erachten.

Die resultierende GSC-Struktur wurde in [Her04] mit einer Echokompensation in unterschiedlichen Varianten als Gesamtsystem untersucht. In [HBNK07] sind weitere Robustheitsaspekte bezüglich der Adaption beschrieben, speziell für den Fall von *Double-Talk*-Situationen.

Besonders wichtig ist hier noch abschließend zu erwähnen, dass bei einem permanent aktiven starken Störgeräuschfeld der GSC mit ASC *Blocking Matrix* zu starken Sprachsignalverzerrungen und einer schlechten Störgeräuschreduktion führen kann [Krü07]. Dies ist offensichtlich, da zum Erreichen der optimalen Koeffizienten in Gl. (8.35) bei der Adaptionsregel Gl. (8.42) in dem Referenzsignal $Y_{\mathrm{FBF}\ m}(\Omega_k)$ nur Sprachkomponenten vorhanden sein dürfen. Für die Problemstellung in dieser Arbeit dient dieses Verfahren also lediglich als Referenzverfahren, welches unter optimalen Bedingungen adaptiert wird.

8.2.4 Neuartige Bestimmung der Blocking Matrix

Wie in Abschnitt 5.2 gezeigt wurde, ist mittels der adaptiven Eigenwertzerlegung eine gute Schätzung für den dominanten Eigenvektor

$$\mathbf{v}_1(\Omega) = \zeta(\Omega)\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}(\Omega) \tag{8.44}$$

möglich. Die optimale Lösung Gl. (8.44) kann nun von links mit $\mathbf{\Phi}_{\mathbf{NN}}(\Omega)$ multipliziert werden

$$\tilde{\mathbf{H}}(\Omega) = \mathbf{\Phi}_{\mathbf{NN}}(\Omega)\mathbf{v}_1(\Omega) \tag{8.45}$$

um die resultierenden Funktionen $\tilde{\mathbf{H}}(\Omega) = (\tilde{H}_1(\Omega), \tilde{H}_2(\Omega), \ldots, \tilde{H}_M(\Omega))^T$ direkt in Gl. (8.20) oder Gl. (8.21) einzusetzen. Der noch verbleibende skalare Faktor $\zeta(\Omega)$ zwischen $\tilde{H}_i(\Omega)$ und $H_i(\Omega)$ spielt dabei keine Rolle, da die Normierung in Gl. (8.20) bzw. Gl. (8.21) dafür sorgt, dass dieser herausfällt. Die so ermittelte *Blocking Matrix* soll mit GTFRBM bezeichnet werden, in Anlehnung an die TFRBM, allerdings hier berechnet mit Hilfe des GEV.

Eine andere Variante [WKHU08] ergibt sich auf der Grundlage der ASCBM aus dem vorherigen Abschnitt. Denn wie in dem Kapitel 4 gezeigt wurde, kann mittels des Filtervektors $\mathbf{v}_1(\Omega)$ – abgesehen von der Skalierung – ein statistisch optimales *Beamforming* erreicht werden. Daher ist das so gefilterte Eingangssignal als optimales Sprachreferenzsignal anzusehen

$$\mathbf{F}_{\mathrm{ref}}(\Omega) = \mathbf{v}_1(\Omega) \tag{8.46}$$

$$Y_{\mathrm{ref}}(\Omega) = \mathbf{v}_1^H(\Omega)\mathbf{X}(\Omega). \tag{8.47}$$

Folgt man dem Ansatz Gl. (8.32), bei dem zwischen dem Referenzsignal und den Eingangssignalen Filter eingefügt werden, so ergeben sich diese durch die Bedingung

$$E\left\{\left(\mathbf{X}(\Omega) - \boldsymbol{\mathcal{G}}(\Omega)Y_{\mathrm{ref}}(\Omega)\right)Y_{\mathrm{ref}}^*(\Omega)\right\}\Big|_{\mathbf{X}(\Omega)=\mathbf{S}(\Omega)+\mathbf{N}(\Omega)} \stackrel{!}{=} 0, \tag{8.48}$$

wobei nun in Gl. (8.48) ein gestörtes Sprachsignal am Eingang zugelassen wird. Das optimale Ergebnis kann durch Ausnutzung der Eigenwertgleichung $\mathbf{\Phi}_{\mathbf{XX}}(\Omega)\mathbf{v}_1(\Omega) = \lambda_1(\Omega)\mathbf{\Phi}_{\mathbf{NN}}(\Omega)\mathbf{v}_1(\Omega)$ angegeben werden zu

$$\boldsymbol{\mathcal{G}}_{\mathrm{opt}}(\Omega) = \frac{\mathbf{\Phi}_{\mathbf{NN}}(\Omega)\mathbf{v}_1(\Omega)}{\mathbf{v}_1^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}(\Omega)\mathbf{v}_1(\Omega)}. \tag{8.49}$$

Da in Gl. (8.49) alle Größen als bekannt angenommen werden, ist keine weitere Adaption wie im Abschnitt 8.2.3 notwendig. Es lässt sich also direkt die GEV *Blocking Matrix* (GEVBM) angeben

$$\mathbf{B}_{\mathrm{GEV}}^H(\Omega) = \mathbf{I} - \frac{\mathbf{\Phi}_{\mathbf{NN}}(\Omega)\mathbf{v}_1(\Omega)\mathbf{v}_1^H(\Omega)}{\mathbf{v}_1^H(\Omega)\mathbf{\Phi}_{\mathbf{NN}}(\Omega)\mathbf{v}_1(\Omega)}, \tag{8.50}$$

wobei der Index "GEV" in Gl. (8.50) auf die Bestimmung mittels des dominanten Eigenvektors hinweist. Selbstverständlich kann Gl. (8.50) mit Gl. (8.45) und Gl. (8.46) in eine zu Gl. (8.39) äquivalente Form umgewandelt werden.

Implementierung der GTFRBM und GEVBM

Die blockorientierte adaptive Bestimmung von $\mathbf{B}_{\mathrm{GEV}}^H(\Omega)$ nach Gl. (8.50) für die diskreten Spektralkomponenten Ω_k erfolgt im Wesentlichen durch die Bestimmung des dominanten Eigenvektors $\mathbf{v}_1(\Omega_k)$ mit Hilfe des Algorithmus 5 (A-PM-GG). Dafür wird zunächst in Sprachpausen durch exponentielle Glättung die Matrix $\hat{\mathbf{\Phi}}_{\mathbf{NN}}^{-1}(\Omega_k)$ bestimmt und während Sprachaktivität $\mathbf{v}_1(\Omega_k)$ sowie die gleichgewichtete Schätzung von $\hat{\mathbf{\Phi}}_{\mathbf{XX}}(\Omega_k)$ aktualisiert (siehe 5.2.1).

Danach erfolgt die Umformung gemäß Gl. (8.50). Für die letztendliche Filterung der Mikrophonsignale ist nun wieder auf die Vermeidung von zyklischen Effekten zu achten. Daher wird ein Verfahren äquivalent zu dem Vorgehen bei der TFRBM eingesetzt. Es werden also die $L = 2B$ Filterkoeffizienten der *Blocking Matrix* zunächst wieder in den Zeitbereich transformiert. Hier sind B Koeffizienten herauszuschneiden und mit Nullen auf die doppelte Länge aufzufüllen. Nach einer erneuten Fourier-Transformation liegen die L Filterkoeffizienten zur Filterung vor. Für den Fall der GTFRBM ist prinzipiell das gleiche Vorgehen anwendbar.

8.3 Fixed Beamformer

Um die GSC-Struktur nach Bild 8.1 zu realisieren ist noch ein geeigneter *Fixed Beamformer* notwendig. In dieser Arbeit werden hierfür zwei Varianten vorgeschlagen. Zum einen ist dies die einfachste Methode mittels DSB und zum anderen ein "blindes" Verfahren basierend auf der Schätzung der Übertragungsfunktionen mittels adaptiver Eigenwertzerlegung.

8.3.1 DSB als FBF

Für den Aufbau eines DSBs sind zwei Komponenten notwendig. Zuerst ist die Sprecherrichtung zu bestimmen und als nächstes sind die jeweiligen Laufzeitunterschiede des direkten Pfades zwischen der Quelle und den Mikrophonen auszugleichen. Der Vorteil bei diesem FBF ist eine unverzerrte Übertragung des Sprachsignals. Der Nachteil ist jedoch die Notwendigkeit einer expliziten Bestimmung der Sprecherrichtung. Dadurch ergibt sich natürlich eine gewisse Einschränkung des angeführten Vorteils, da nur dann ein unverzerrtes Nutzsignal am Ausgang erreicht wird, wenn die DOA auch korrekt ermittelt wird. Weiterhin gilt diese Einschränkung ebenfalls für den Aspekt einer optimalen Realisierung der Laufzeitkompensation.

Wie in Kapitel 7 gezeigt wurde, ist mittels der Methode der Abtastung der Richtcharakteristik unter Verwendung des generalisierten dominanten Eigenvektors $\hat{\mathbf{v}}_1(\Omega)$ in Gl. (7.15) eine sehr gute Schätzung der Sprecherrichtung möglich. Und zwar auch in Umgebungen mit gerichteten Störschallquellen. Daher soll dieses Verfahren zur Bestimmung der DOA in der GSC-Struktur Verwendung finden.

Zur Kompensation der Laufzeitunterschiede sind in [LVKL96] verschiedene Verfahren zur Realisierung von Verzögerungen kleiner als die Abtastzeit zusammengestellt. Ein Problem stellt dabei insbesondere die frequenzunabhängige Signaldämpfung dar, die je nach gewähltem Verfahren stark von der umzusetzenden Verzögerung abhängt. Hinzu kommt noch der nicht zu unterschätzende Rechenaufwand für die fortlaufende Berechnung der Interpolationsfilter in Abhängigkeit der ermittelten DOA. Daher soll hier eine gänzlich andere Methode zur Laufzeitkompensation vorgeschlagen werden, die sich bei der Realisation des Gesamtsystem als sehr effizient erwiesen hat.

Die Untersuchungen zur Sprachverzerrung durch eine fehlerhafte Laufzeitkompensation in Kapitel 3.5 haben gezeigt, dass eine geringe Abweichung zwischen Zielrichtung des *Arrays* und tatsächlicher Sprecherrichtung als durchaus tolerierbar einzustufen ist. Daher ist es sinnvoll, für eine konkrete geometrische Anordnung *a priori* Interpolationsfilter für ein bestimmtes Raster von Zielrichtungen zu berechnen und in einer Datenbank abzulegen. Diese Filterkoeffizienten müssen dann zur Laufzeit der Software für die ermittelten DOAs nur noch aus der Datenbank ausgelesen, aber nicht mehr berechnet werden. Eine Winkelauflösung von

$\Delta\theta_t = 4^\circ$ erscheint hierbei ausreichend und ergibt somit $2N+1 = 45$ mögliche Zielrichtungen

$$\theta_{t\,\nu} = \nu\Delta\theta_t, \qquad \nu = -N, \ldots, N. \tag{8.51}$$

Die Filterkoeffizienten $\mathbf{F}_{\mathrm{PCA}\,\nu}(\Omega)$ für die Richtungen $\theta_{t\,\nu}$ werden wie folgt berechnet. In einer simulierten Umgebung mit Freifeldausbreitung ($T_{60} = 0\,\mathrm{s}$) wird jeweils an diesen Zielrichtungen eine Quellen platziert, welche weißes Rauschen emittiert. Mittels PCA *Beamforming* werden dann die optimalen Filterkoeffizienten berechnet, wie in Abschnitt 5.1.4 beschrieben ist. Somit ist gewährleistet, dass ein optimaler Laufzeitausgleich gegeben eine bestimmte Filterlänge realisiert wird.

Für die eigentliche Filterung zur Laufzeit ist dann schließlich der Koeffizientensatz zu wählen, der zu dem Index der Richtung gehört, für die gilt

$$\hat{\nu} = \underset{\nu}{\mathrm{argmin}} |\hat{\theta}_s - \theta_{t\,\nu}|, \tag{8.52}$$

wobei $\hat{\theta}_s$ die geschätzte Sprecherrichtung ist. Das Ausgangssignal ist somit gegeben durch

$$Y_{\mathrm{FBF}}(\Omega) = \mathbf{F}^H_{\mathrm{PCA}\,\hat{\nu}}(\Omega)\mathbf{X}(\Omega). \tag{8.53}$$

Bei der blockorientierten Implementierung ist die Filterung Gl. (8.53) wieder mittels *Overlap-Save*-Methode [Shy92] für diskrete Spektralkomponenten Ω_k umzusetzen, wobei natürlich aufgrund der zeitabhängigen Schätzung der Sprecherrichtung auch die Wahl der Filterkoeffizienten von Block zu Block unterschiedlich sein kann.

8.3.2 Matched Filter als FBF

Die explizite Bestimmung der Sprecherrichtung kann vermieden werden, wenn der dominante Eigenvektor Gl. (8.44) in einer nachverarbeiteten Version zur Filterung hergenommen wird. Dazu sind zunächst wieder die skalierten Raumübertragungsfunktionen $\tilde{\mathbf{H}}(\Omega) = \mathbf{\Phi_{NN}}(\Omega)\mathbf{v}_1(\Omega)$ zu bestimmen, welche dann entsprechend der BAN-Methode aus Abschnitt 6.4.1 normiert werden (vgl. Gl. (6.18) bzw. Gl. (6.19))

$$\mathbf{F}_{\mathrm{MF}}(\Omega) = \frac{1}{\sqrt{M}} \frac{\tilde{\mathbf{H}}(\Omega)}{||\tilde{\mathbf{H}}(\Omega)||}. \tag{8.54}$$

Für die Filterkoeffizienten in Gl. (8.54) wurde der Index "MF" als Kennzeichnung für das *Matched Filter* verwendet. Denn obschon die Koeffizienten äquivalent zum PCA *Beamforming* und ines damit verbundenen *Matched Filters* sind, basiert die Bestimmung des Eigenvektors nicht auf dem speziellen, sondern dem allgemeinen Eigenwertproblem.

Gl. (8.54) basiert auf der Näherung $||\mathbf{H}(\Omega)|| \approx \sqrt{M}$, wodurch dann folglich auch nur näherungsweise ein unverzerrtes Sprachsignal am Ausgang des FBFs zu erwarten ist:

$$\begin{aligned}
Y_{\mathrm{FBF}}(\Omega) &= \mathbf{F}^H_{\mathrm{MF}}(\Omega)\mathbf{X}(\Omega) && (8.55)\\
&= \mathbf{F}^H_{\mathrm{MF}}(\Omega)\big(S_c(\Omega)\mathbf{H}(\Omega) + \mathbf{N}(\Omega)\big) && (8.56)\\
&= \frac{\zeta(\Omega)}{|\zeta(\Omega)|}\frac{||\mathbf{H}(\Omega)||}{\sqrt{M}} S_c(\Omega) + \mathbf{F}^H_{\mathrm{MF}}(\Omega)\mathbf{N}(\Omega) && (8.57)\\
&\approx \frac{\zeta(\Omega)}{|\zeta(\Omega)|} S_c(\Omega) + \mathbf{F}^H_{\mathrm{MF}}(\Omega)\mathbf{N}(\Omega). && (8.58)
\end{aligned}$$

Bei einer kleinen Nachhallzeit erhält man ein nahezu unverzerrt gefiltertes Signal aus der implizit ermittelten Sprecherrichtung. Für große Nachhallzeiten ist mit einer gering variierenden, frequenzselektiven Dämpfung für die gewünschte Richtung zu rechnen. Dieses Verhalten ist beispielhaft an den Richtcharakteristiken in Bild 8.2 verdeutlicht. Dargestellt sind die *Beampattern* der Filterkoeffizienten nach Gl. (8.54) bei Anwendung des Verfahrens in dem *Szenario-2* für zwei unterschiedliche Nachhallzeiten.

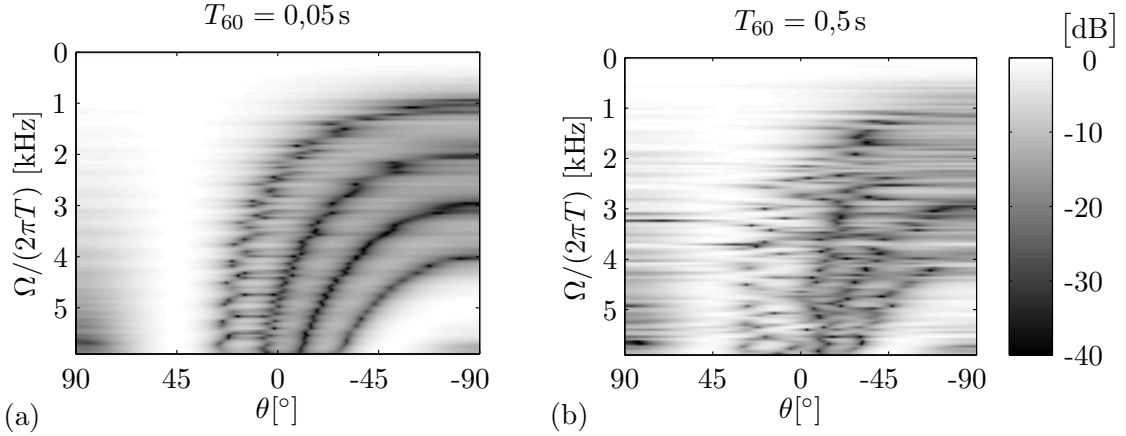

Bild 8.2: Richtcharakteristiken der Koeffizienten des *Matched Filters* als FBF für die Nachhallzeiten von $T_{60} = 0{,}05\,\mathrm{s}$ und $T_{60} = 0{,}5\,\mathrm{s}$. Die Sprecherrichtung beträgt $\theta_s = 45\,^\circ$ und das gerichtete Tiefpassrauschen hat eine Einfallsrichtung von $\theta_n = -20\,^\circ$.

An dem *Beampattern* für die Nachhallzeit $T_{60} = 0{,}05\,\mathrm{s}$ in Bild 8.2 (a) ist sehr gut das zu einem DSB äquivalente Verhalten zu erkennen. Für die Nachhallzeit $T_{60} = 0{,}5\,\mathrm{s}$ in Bild 8.2 (b) ist die für einen *Matched Filter Beamformer* typische Charakteristik wiederzufinden. Es bildet sich für die Sprecherrichtung $\theta_s = 45\,^\circ$ nur näherungsweise für alle Frequenzen die gleiche Dämpfung aus, da bei der Ermittlung des dominanten Eigenvektors noch frühe Reflexionen berücksichtigt werden. Bei subjektiven Hörtest hat sich dieses Verhalten jedoch als nicht signifikant erwiesen. Da also der *Matched Filter Beamformer* den Vorteil einer blinden Arbeitsweise aufweist, ist hierin eine sehr gute Alternative zur Realisierung als FBF in einer GSC-Struktur zu sehen.

8.4 Experimentelle Untersuchungen

Im Folgenden sollen Ergebnisse zu den experimentellen Untersuchungen der GSC-Strukturen mit den beiden unterschiedlichen *Fixed Beamformern* präsentiert werden. Zunächst ist dies die Realisierung des *Fixed Beamformers* als DSB und im Anschluss die Variante mittels eines *Matched Filters*. Grundsätzlich gilt hier wieder bei den adaptiven Filtern, dass alle Messungen mit konvergierten Koeffizienten vorgenommen wurden. Dies betrifft die adaptiven *Blocking-Matrix*-Varianten, die *Adaptive Noise Cancellation* und den *Matched Filter Beamformer*. Der DSB ist optimal realisiert mit dem *a priori* Wissen über die Sprecherrichtung und einer exakten Laufzeitkompensation.

Für die Ergebnisse in den nachfolgenden Diagrammen sollen folgende Abkürzungen definiert sein:

- *Generalized Sidelobe Canceller* mit *Delay-and-Sum-Beamformer* als *Fixed Beamformer* und verschiedenen Varianten der *Blocking Matrix*

  - DOR: *Delay Only Ratio Blocking Matrix* gemäß Gl. (8.24) durch die paarweise Subtraktion zeitangepasster Mikrophonsignale nach Griffiths und Jim [GJ82]
  - TFR: *Transfer Function Ratio Blocking Matrix* Gl. (8.20) mit der Bestimmung des Verhältnisses der Übertragungsfunktionen nach Gannot et al. [GBW01] mit Gl. (8.30)
  - ASC: *Adaptive Speech Cancellation Blocking Matrix* mit Hilfe adaptiver Filter und NLMS-Adaption mit Gl. (8.42) und Gl. (8.43), wobei das reine Sprachsignal am DSB-Ausgang als Referenzsignal[3] dient
  - GTFR: *Generalized Eigenvector Transfer Function Ratio Blocking Matrix* basierend auf der BM Gl. (8.20), wobei die Übertragungsfunktionen mittels Algorithmus 5 (A-PM-GG) bestimmt werden
  - GEV: *Generalized Eigenvector Blocking Matrix* entsprechend der neuartigen Form in Gl. (8.50)

- *Generalized Sidelobe Canceller* mit *Matched Filter* nach Gl. (8.54) als *Fixed Beamformer* und beide Varianten der *Blocking Matrix* basierend auf dem dominanten Eigenvektor
  - MF-GTFR: *Matched Filter* FBF und *Generalized Eigenvector Transfer Function Ratio Blocking Matrix*
  - MF-GEV: *Matched Filter* FBF und *Generalized Eigenvector Blocking Matrix*

Alle adaptiven Filter sind im Frequenzbereich unter Anwendung der blockorientierten *Overlap-Save*-Methode realisiert worden. Bis auf eine explizit gekennzeichnete Ausnahme wurden für das Verfahren nach Gannot und die eigenvektorbasierten Methoden eine Filterlänge von $B = 256$ Koeffizienten gewählt. Das *Matched Filter* FBF ist jedoch mit einer Filterlänge von 128 für jeden Mikrophonpfad implementiert worden. Dafür kann sehr effizient aus dem adaptiv berechneten dominanten Eigenvektor in der entsprechenden *Blocking Matrix* jede zweite Frequenzkomponente entnommen werden. Die Motivation für eine geringere Filterlänge im FBF ist in Abschnitt 6.4.2 zu finden.

Die mehrkanalige *Adaptive Noise Cancellation* ist mit einer Filterlänge von 1024 pro Pfad realisiert, wobei die Filterkoeffizienten gemäß der normalisierten LMS-Adaptionsregel Gl. (8.15) und Gl. (8.16) bestimmt wurden.

Grundsätzlich wird bei allen Simulationen den Eingangsdaten wieder jeweils weißes, räumlich unkorreliertes Rauschen mit einem SNR von 25 dB hinzugefügt. Desweiteren werden die jeweiligen räumlich korrelierten Störsignale mit einem SNR von 5 dB additiv überlagert.

8.4.1 Generalized Sidelobe Canceller mit DSB

Gemäß Gl. (8.7) sollten die Störgeräuschreferenzsignale im Idealfall keinen Sprachanteil mehr enthalten. Dies ist natürlich insbesondere für steigende Nachhallzeiten aufgrund der begrenzten Filterlänge in der *Blocking Matrix* und den jeweiligen Schätzfehlern der verwendeten Verfahren nur bedingt zu erzielen. Um das Vermögen der Sprachblockierung (engl. *Blocking*

[3] Es soll nochmal darauf hingewiesen werden, dass für die ASCBM in der Praxis nicht das reine Sprachsignal am DSB-Ausgang beobachtet werden kann und daher diese Anordnung nur zu Vergleichszwecken verwendet wird.

Ability, BA) einer *Blocking Matrix* zu messen, soll im Zeitbereich die Dämpfung des Sprachsignals relativ zur Störung vom Eingang zum Ausgang für die betrachtete BM wie folgt bestimmt werden

$$\mathrm{BA} := 10 \cdot \left[\log_{10} \left(\frac{\sum_{i=1}^{M} \sum_{n \in T_s} u_{s,i}^2(n)}{\sum_{i=1}^{M} \sum_{n \in T_s} u_{n,i}^2(n)} \right) - \log_{10} \left(\frac{\sum_{i=1}^{M} \sum_{n \in T_s} x_{s,i}^2(n)}{\sum_{i=1}^{M} \sum_{n \in T_s} x_{n,i}^2(n)} \right) \right] \mathrm{dB}. \tag{8.59}$$

Es wird also in Gl. (8.59) von dem mittleren SNR in den M Störgeräuschreferenzsignalen im logarithmischen Bereich das mittlere SNR in den Mikrophonsignalen, unter Beachtung der Menge der Zeitindizes T_s welche Sprache beinhalten, subtrahiert. $u_{s,i}(n)$ bezeichnet den Sprachanteil im i-ten Störgeräuschreferenzsignal und $u_{n,i}(n)$ entsprechend den Rauschanteil.

In Bild 8.3 (a) ist die *Blocking Ability* für das *Szenario-2* und in Bild 8.3 (b) der SNR-Gewinn dargestellt. Wie erwartet wird für den idealen Fall mit der ASCBM die größte Dämpfung des Sprachsignals erzielt. Die BA der ASCBM setzt sich insbesondere bei der Freifeldausbreitung deutlich von der BA der anderen Verfahren ab. Obschon für $T_{60} = 0\,\mathrm{s}$ gerade die korrekte Randbedingung für den Einsatz der DORBM eingehalten wird, sind doch minimale Fehler bezüglich der Zeitanpassung aufgrund der Annahme einer planar auf das *Array* einfallenden Schallwelle vorhanden (trotz bekannter Sprecherrichtung). Zusätzlich sind Pegeldifferenzen zwischen den Sensorsignalen nicht kompensiert. Beiden Effekten kann jedoch mit den adaptiven Filtern in der ASCBM optimal begegnet werden. Mit steigender Nachhallzeit steigt auch der Sprachanteil in den Störgeräuschreferenzsignalen für alle BM-Varianten. Hier liegen die Werte der *Blocking Ability* der GTFRBM und GEVBM im Bereich zwischen den Ergebnissen für die ASCBM und die DORBM. Hingegen unterscheiden sich die Verläufe der BA für die TFRBM und DOR nicht wesentlich voneinander.

Der SNR-Gewinn in Bild 8.3 (b) zeigt, dass die Verläufe für die GSC-Strukturen mit GEVBM und GTFRBM dem optimalen Verlauf bei der Realisierung mit der ASCBM sehr nahe kommen. Leider liefert hier die Struktur mit TFRBM nicht die erwartete Leistungsfähigkeit. Der SNR-Gewinn liegt signifikant unter den Ergebnissen der anderen Verfahren und ist nur für größere Nachhallzeiten ähnlich zu dem SNR-Gewinn mit der DORBM. Maßgeblich ist hierfür eine schlechte Unterdrückung der unteren Frequenzkomponenten der Sprache durch die TFRBM, welche insbesondere bei geringen Nachhallzeiten ins Gewicht fällt. Damit verbunden ist eine schlechte Rauschunterdrückung des GSCs im unteren Frequenzbereich und eine generelle Anhebung des gefilterten Signals für diese Frequenzen. Dieses Verhalten soll durch eine genauere Betrachtung des reinen Sprachsignals am GSC-Ausgang verdeutlicht werden. Hierfür wird das Verhältnis der spektralen Leistungsdichte der reinen Sprachsignale vor und nach der Subtraktion über den *Noise-Cancellation*-Pfad gebildet:

$$\delta_{\mathrm{LDS}}(\Omega) = \left. \frac{\hat{\phi}^{(\mathrm{GG})}_{Y_{\mathrm{GSC}} Y_{\mathrm{GSC}}}(\Omega)}{\hat{\phi}^{(\mathrm{GG})}_{Y_{\mathrm{FBF}} Y_{\mathrm{FBF}}}(\Omega)} \right|_{\mathbf{X}(\Omega)=\mathbf{S}(\Omega)} . \tag{8.60}$$

In Gl. (8.60) beschreibt $\hat{\phi}^{(\mathrm{GG})}_{Y_{\mathrm{FBF}} Y_{\mathrm{FBF}}}(\Omega)$ die über die gesamte Sprachäußerung gleichgewichtet gemittelte spektrale Leistungsdichte nach dem FBF und $\hat{\phi}^{(\mathrm{GG})}_{Y_{\mathrm{GSC}} Y_{\mathrm{GSC}}}(\Omega)$ entsprechend das gemittelte Leistungsdichtespektrum nach dem GSC jeweils für das reine Sprachsignal. Die Abweichung $\delta_{\mathrm{LDS}}(\Omega)$ ist in Bild 8.4 (a) gemittelt über alle 10 Sprachbeispiele für die GSC-Strukturen mit TFRBM, GTFRBM und GEVBM exemplarisch für eine Nachhallzeit von

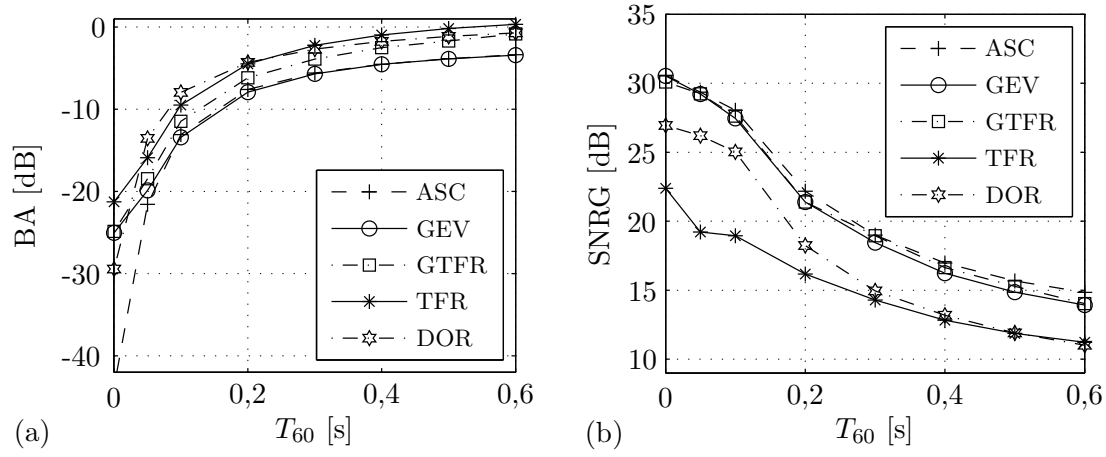

Bild 8.3: *Blocking Ability* in (a) und SNR-Gewinn in (b) für eine Sprecherrichtung von $\theta_s = 45\,^\circ$ und eine Störquelle bei $\theta_n = -20\,^\circ$.

$T_{60} = 0{,}1\,\mathrm{s}$ dargestellt und in Bild 8.4 (b) für die GSC-Strukturen mit DORBM und ASCBM. Für den GSC mit TFRBM ist eine auffällige Anhebung der Spektralkomponenten bis ca. 500 Hz zu erkennen, welche auch bereits in [GBW04] Erwähnung fand. Mit der GTFRBM und GEVBM erfolgt hingegen eine leichte Dämpfung der unteren Frequenzkomponenten. Abgesehen von dem GSC mit DORBM ergibt sich für die anderen Methoden eine Dämpfung des Signals für die höchsten Frequenzen, da hier nahezu kein Sprachsignal vorhanden ist.

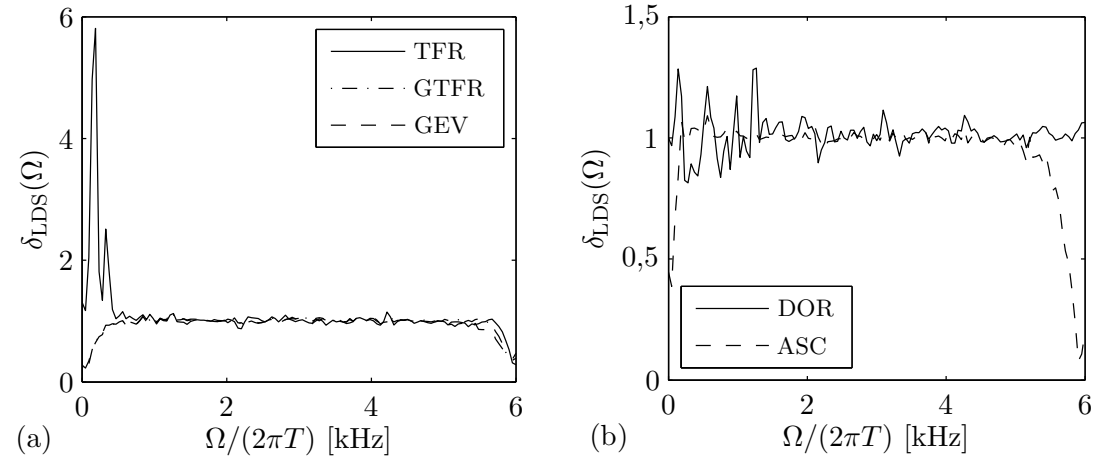

Bild 8.4: LDS-Verhältnisse nach Gl. (8.60) für eine Sprecherrichtung von $\theta_s = 45\,^\circ$ und eine Störquelle bei $\theta_n = -20\,^\circ$ für eine Nachhallzeit von $T_{60} = 0{,}1\,\mathrm{s}$.

Daher soll nun die Varianz der spektralen Abweichung für die Spektralkomponenten korrespondierend zu dem Frequenzbereich zwischen ca. 0,5 kHz und 5 kHz untersucht werden: $\sigma^2_{\mathrm{LDS}} := \mathrm{var}\{\delta_{\mathrm{LDS}}(\Omega)\}$. Eine Varianz von Null besagt, dass alle Frequenzkomponenten gleich stark gedämpft bzw. verstärkt werden und sich somit lediglich eine Lautstärkeänderung ergeben kann. Große Werte für die Varianz bedeuten hingegen, dass die verschiedenen Frequenzkomponenten unterschiedlich stark gedämpft oder verstärkt wurden, was folglich zu einer Sprachverzerrung führt. Die Varianz σ^2_{LDS} wird wieder über alle Beispieläußerungen gemittelt und über der Nachhallzeit betrachtet. Für das *Szenario-2* sind die Ergebnisse in Bild 8.5 (a) dargestellt. Die Varianz σ^2_{LDS} in Bild 8.5 zeigt für alle Verfahren geringe Werte

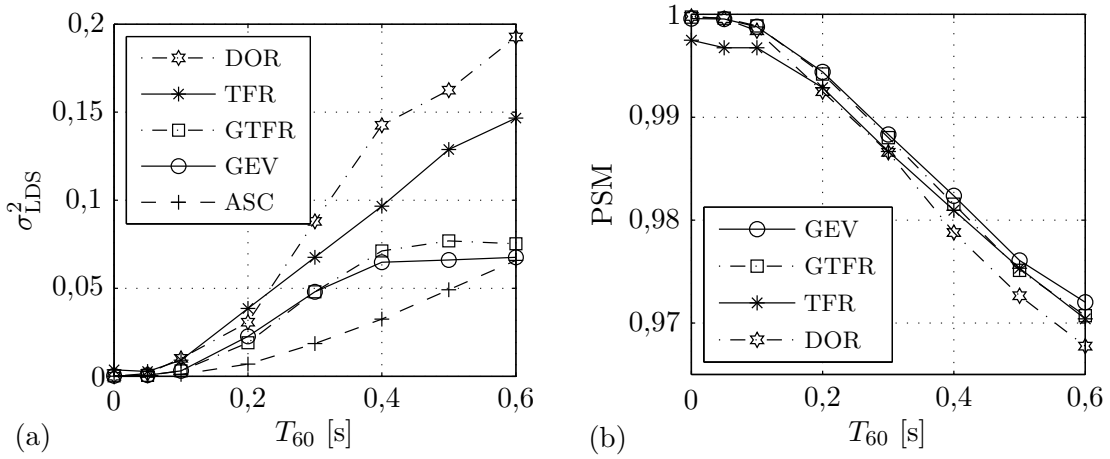

Bild 8.5: In (a) Varianz der Verhältnisse der spektralen Leistungsdichte des GSC-Ausgangssignals zu der des FBFs -Ausgangssignal und in (b) das perzeptive Qualitätsmaß für eine Sprecherrichtung von $\theta_s = 45\,^\circ$ und einer Störquelle bei $\theta_n = -20\,^\circ$.

für kleine Nachhallzeiten. Mit steigendem T_{60} weist dann der GSC mit DORBM die höchsten Werte für die Varianz auf gefolgt von der TFR-Methode. Die beiden Realisierungen mit dem dominanten Eigenvektor in der GTFRBM und der GEVBM weisen nur geringe Unterschiede zueinander auf. Die geringste Varianz ergibt sich schließlich für das Referenzsystem mit ASCBM. Die Ergebnisse der Varianzmessung decken sich prinzipiell mit den Ergebnissen der perzeptiven Sprachqualitätsmessung, welche in Bild 8.5 (b) zu sehen sind. Dabei sind nun wieder alle Spektralkomponenten beteiligt und pro Nachhallzeit ist der Mittelwert der PSM-Werte der 10 verwendeten Beispielsätze abgebildet. Als Referenzsignal wurde jeweils das reine Sprachsignal des GSC-Referenzsystems mit ASCBM verwendet. Die auffällig geringeren PSM-Werte für das TFRBM System bei kleinen Nachhallzeiten sind wieder durch die Tiefenanhebung zu erklären.

Als nächstes folgen Ergebnisse zu den gleichen Messungen wie zuvor, jedoch für das diffuse Störschallfeld bei weiterhin einer Sprecherrichtung von $\theta_s = 45\,^\circ$. Die *Blocking Ability* und der SNR-Gewinn für diese Anordnung sind in Bild 8.6 dargestellt. Die *Blocking Ability* der DORBM und ASCBM sind nahezu identisch zu den entsprechenden Verläufen in Bild 8.3, jedoch sind die Ergebnisse für die GEVBM geringfügig schlechter und für die TFRBM geringfügig besser. Der SNR-Gewinn für den GSC mit TFRBM liegt nun auch leicht über der Methode mit DORBM, wobei weiterhin – abgesehen von dem Referenzsystem – der GSC mit GEVBM die größte Rauschunterdrückung liefert.

Die Abweichung der spektralen Leistungsdichte für das System mit TFRBM hat sich im unteren Frequenzbereich deutlich verringert, was beispielhaft an Bild 8.7 zu sehen ist. Daher ist auch die Varianz σ^2_{LDS} dieser Realisierung ähnlich zu denen der GSCs mit GTFRBM und GEVBM. Auffällig an den Verläufen der Varianz in Bild 8.8 sind die relativ geringen Werte für den GSC mit DORBM. Dieses Verhalten liegt an der Tatsache, dass hier insgesamt nur recht geringe Signalanteile über das *Sidelobe Cancellation* eliminiert werden. Dies ist an dem kleinen SNR-Gewinn zu erkennen. Daher wird auch das Sprachsignal am Ausgang des FBFs nur geringfügig angegriffen, was auch an dem PSM-Verlauf in Bild 8.7 wiederzufinden ist. Ebenfalls kann auch die relativ gute Sprachqualität der Struktur mit TFRBM für den Fall

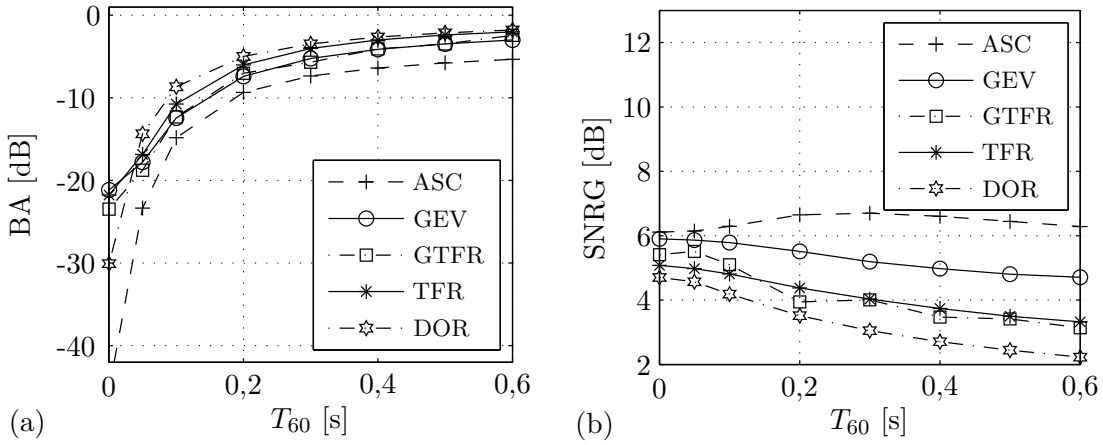

Bild 8.6: *Blocking Ability* in (a) und SNR-Gewinn in (b) für eine Sprecherrichtung von $\theta_s = 45\,^\circ$ und diffusen Störschall.

des diffusen Störschallfelds an den PSM-Werten abgelesen werden.

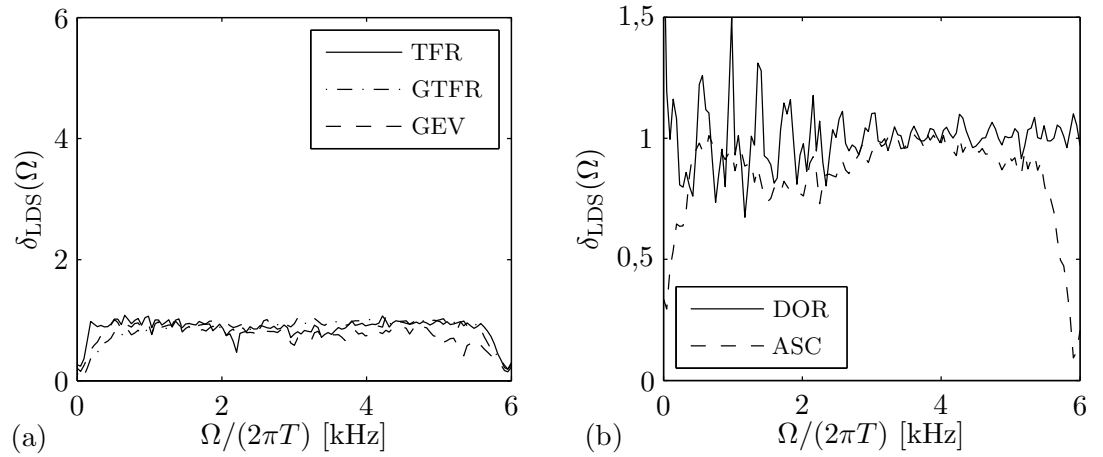

Bild 8.7: LDS-Verhältnisse nach Gl. (8.60) für eine Sprecherrichtung von $\theta_s = 45\,^\circ$ und ein diffuses Störschallfeld.

Für eine Sprecherrichtung von $\theta_s = 0\,^\circ$ und einer Störquelle bei $\theta_n = 60\,^\circ$ gemäß *Szenario-3* sind die *Blocking Ability* und der SNR-Gewinn in Bild 8.9 dargestellt. Die Sprachsignalunterdrückung ist insgesamt für alle *Blocking-Matrix*-Realisierungen für das *Szenario-3* größer als für das *Szenario-2*. Bei der TFRBM sind die Werte der BA zwar für geringe Nachhallzeiten schlechter im Vergleich zu den Werten der GTFRBM und GEVBM, aber für höhere Nachhallzeiten durchaus ähnlich zu diesen. Trotzdem sind die SNR-Gewinne für alle Verfahren etwas geringer im Vergleich zu dem *Szenario-2*. Außerdem ist nun der SNR-Gewinn für den GSC mit DORBM sehr ähnlich zu den Methoden mit den eigenvektorbasierten *Blocking-Matrix*-Verfahren. Für das Verfahren mit TFRBM macht sich allerdings wieder die schlechte Rauschunterdrückung in dem unteren Frequenzbereich bemerkbar, insbesondere bei geringen Nachhallzeiten.

Für den GSC mit TFRBM ist bei dem *Szenario-3* eine signifikante Abweichung $\delta_{\mathrm{LDS}}(\Omega)$ der Leistungsdichtespektren beobachtet worden. Dies ist beispielhaft für eine Nachhallzeit

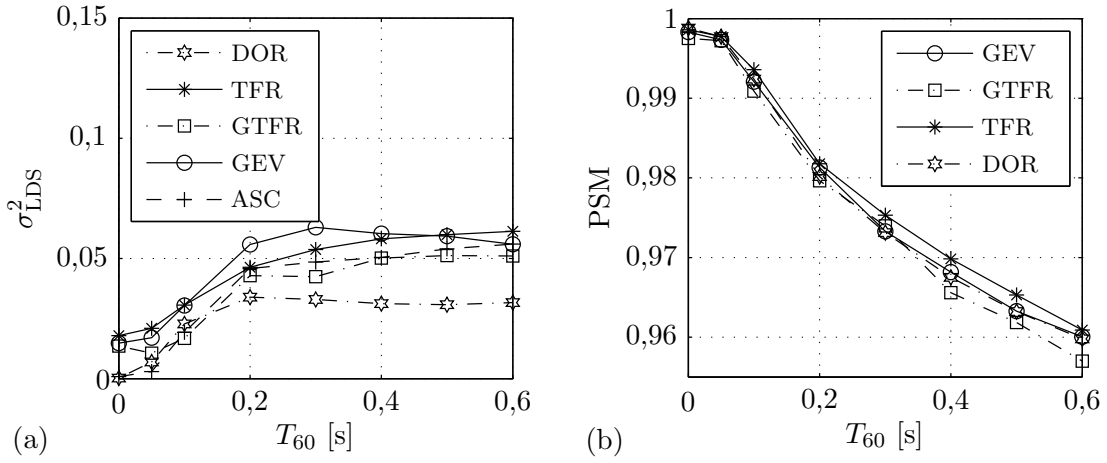

Bild 8.8: In (a) Varianz der Verhältnisse der spektralen Leistungsdichte des GSC-Ausgangssignals zu dem FBF Ausgangssignal und in (b) das perzeptive Qualitätsmaß für eine Sprecherrichtung von $\theta_s = 45\,^\circ$ und einem diffusen Störschallfeld.

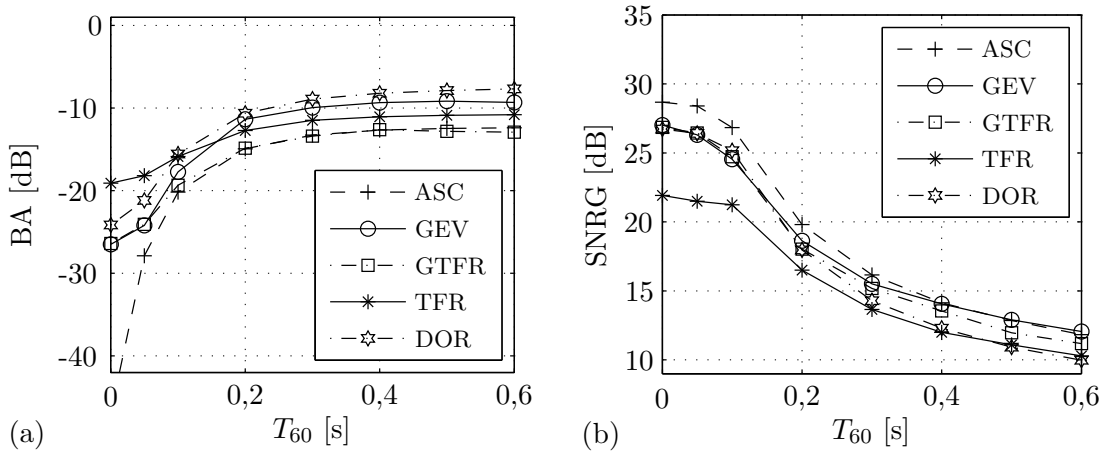

Bild 8.9: *Blocking Ability* in (a) und SNR-Gewinn in (b) für eine Sprecherrichtung von $\theta_s = 0\,^\circ$ und eine Störquelle bei $\theta_n = 60\,^\circ$.

von $T_{60} = 0{,}1\,\mathrm{s}$ in Bild 8.10 zu sehen. Damit verbunden fällt dann auch die Varianz der Abweichungen deutlich höher aus, wie an dem Verlauf in Bild 8.11 (a) zu erkennen ist. Die Verläufe für die Strukturen mit GTFRBM und GEVBM sind ähnlich zu dem Verlauf des Referenzsystems und liegen deutlich unter dem des GSCs mit DORBM. Diese Ergebnisse gehen konform mit der gemessenen perzeptiven Sprachqualität, was an den PSM-Werten in Bild 8.11 (b) abzulesen ist.

Als letztes folgen noch die Ergebnisse für das *Szenario-4*, also für die Anordnung einer Sprachquelle bei $\theta_s = 0\,^\circ$ und zwei Störquellen: eine bei -20 ° und eine bei 60 °. Für dieses Szenario ist nun die Sprachsignalblockierung der GEVBM schlechter als für die anderen adaptiven Verfahren, wie an Bild 8.12 (a) zu sehen ist. Aufgrund der komplizierteren Anordnung ist der SNR-Gewinn für alle GSC-Varianten geringer im Vergleich zum *Szenario-2* und *Szenario-3*. Der Verlauf des SNR-Gewinns des GSCs mit TFRBM ist recht ähnlich zum GSC mit DORBM. Hingegen liefern die eigenvektorbasierten Methoden eine leicht höhere Störgeräuschunterdrückung.

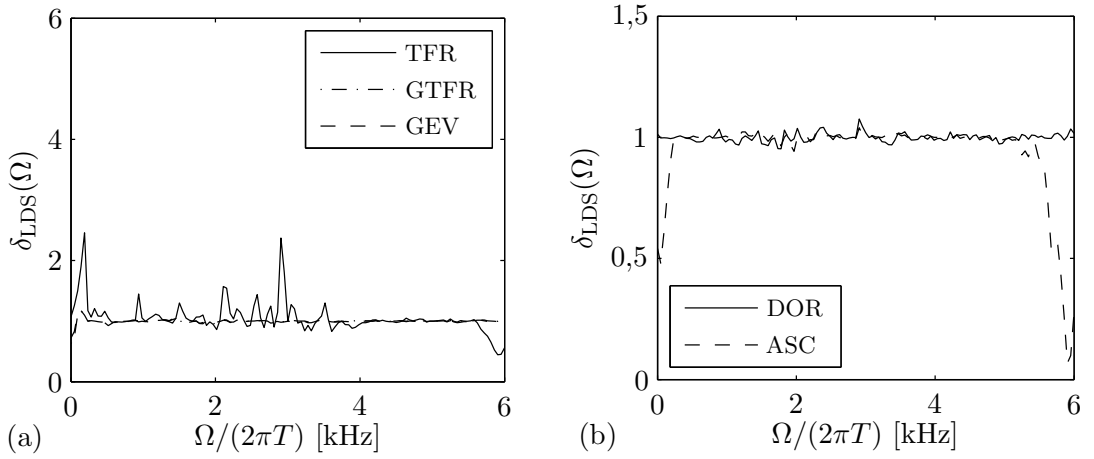

Bild 8.10: LDS-Verhältnisse nach Gl. (8.60) für eine Sprecherrichtung von $\theta_s = 0\,^\circ$ und eine Störquelle bei $\theta_n = 60\,^\circ$.

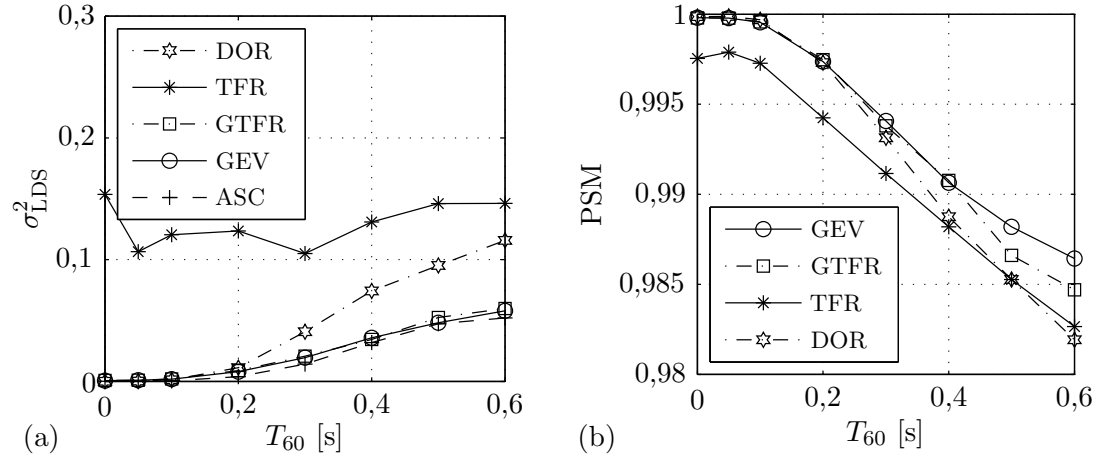

Bild 8.11: In (a) Varianz der Verhältnisse der spektralen Leistungsdichte des GSC-Ausgangssignals zu der des FBFs-Ausgangssignal und in (b) das perzeptive Qualitätsmaß für eine Sprecherrichtung von $\theta_s = 0\,^\circ$ und einer Störquelle bei $\theta_n = 60\,^\circ$.

Die spektrale Abweichung $\delta_{\mathrm{LDS}}(\Omega)$ ist hier für den GSC mit TFRBM bei den tiefen Frequenzen nicht so deutlich ausgeprägt, wie beispielhaft an Bild 8.13 zu erkennen ist. Dennoch sind stärkere Abweichungen über den gesamten Frequenzbereich beobachtet worden als für die Verfahren mit GTFRBM und GEVBM. Für diese zeigt die Varianz σ^2_{LDS} in Bild 8.14 (a) sehr ähnliche Verläufe wie das Referenzsystem. Aber dennoch ist erstaunlicherweise die resultierende Sprachqualität aufgrund eines leichten Hochpass-Charakters geringfügig schlechter im Vergleich zum System mit TFRBM.

Die Simulationsergebnisse für die GSC-Strukturen mit einem DSB als *Fixed Beamformer* können wie folgt zusammengefasst werden:

- Die ASCBM liefert natürlich die besten Resultate, da die Adaption ja mit dem reinen Sprachsignal am DSB-Ausgang erfolgt.
- Die eigenvektorbasierten *Blocking-Matrix*-Methoden GTFRBM und GEVBM unterscheiden sich nur geringfügig. Dennoch liefert die GEVBM aber eine leicht bessere

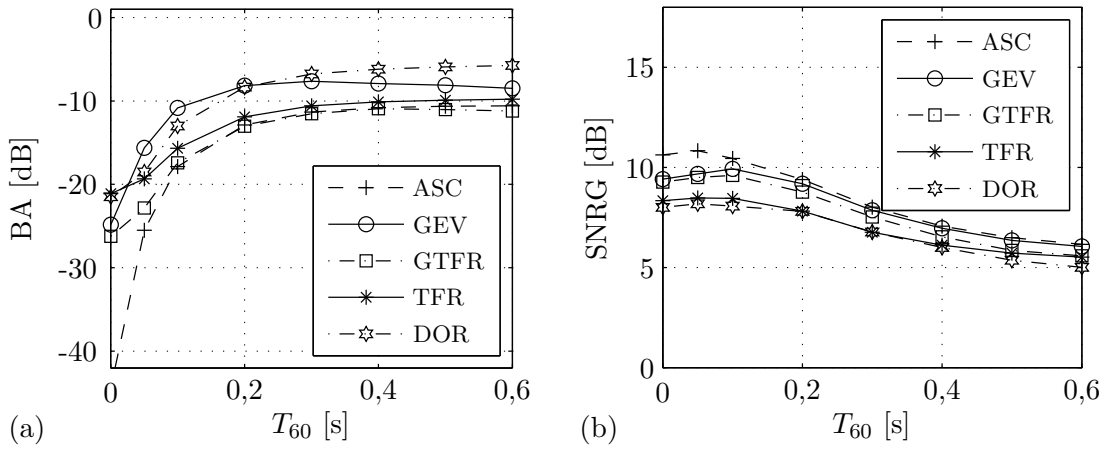

Bild 8.12: *Blocking Ability* in (a) und SNR-Gewinn in (b) für eine Sprecherrichtung von $\theta_s = 0\,^\circ$ und zwei Störquellen: eine bei -20 ° und eine bei 60 °.

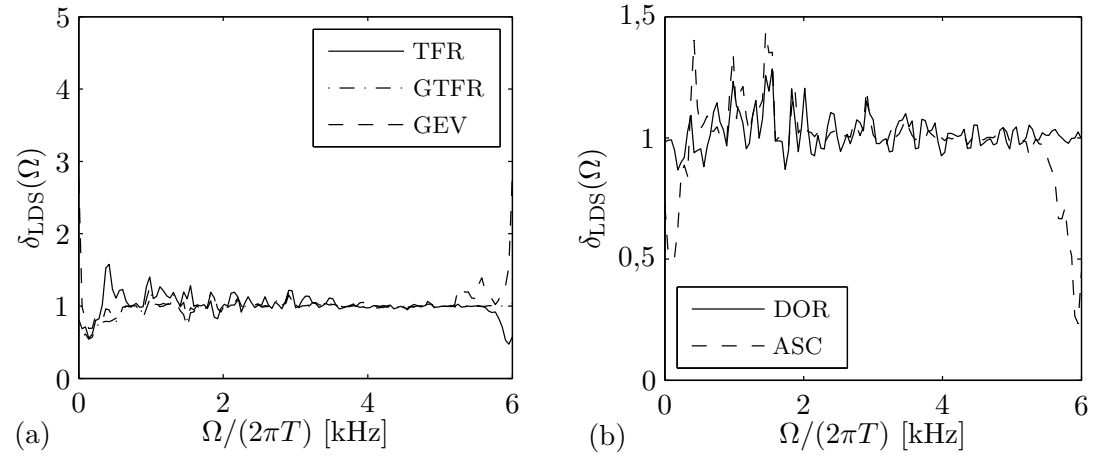

Bild 8.13: LDS-Verhältnisse nach Gl. (8.60) für eine Sprecherrichtung von $\theta_s = 0\,^\circ$ und zwei Störquellen: eine bei -20 ° und eine bei 60 °.

Rauschunterdrückung bei einem tendenziell unverfälschterem Sprachsignal.

- Die nichtadaptive Realisierung der *Blocking Matrix* als DORBM zeigt gute Ergebnisse, die jedoch deutlich unter denen der eigenvektorbasierten Methoden liegen.

- Die Leistungsfähigkeit des GSCs mit TFRBM ist stark abhängig von der konkreten Anordnung. Bei zahlreichen Experimenten hat diese Realisierung bezüglich des SNR-Gewinns und der Sprachqualität schlechtere Ergebnisse erzielt als die konventionelle nichtadaptive Methode. Insbesondere treten hier häufig Probleme im unteren Frequenzbereich auf.

Nach den ausführlichen Betrachtungen der Simulationsergebnisse für unterschiedliche Anordnungen der Schallquellen bleibt die Frage nach der Auswirkung von unterschiedlich gewählten Parametern. Hierzu können folgende Aussagen getroffen werden:

- Eine variierende Anzahl M der verwendeten Mikrophone hat maßgeblichen Einfluss auf die Unterdrückung von räumlich unkorreliertem Rauschen, also das additive Mi-

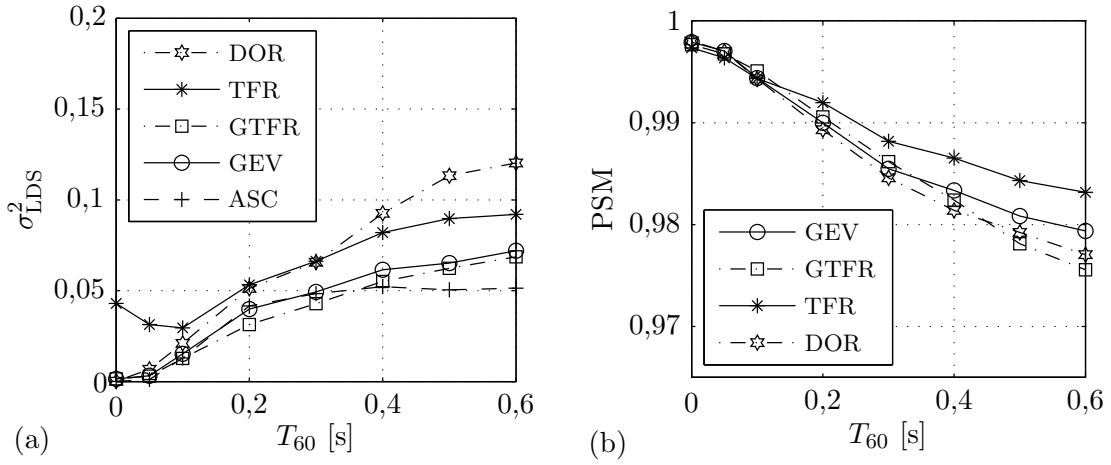

Bild 8.14: In (a) Varianz der Verhältnisse der spektralen Leistungsdichte des GSC-Ausgangssignals zu der des FBFs-Ausgangssignal und in (b) das perzeptive Qualitätsmaß für eine Sprecherrichtung von $\theta_s = 0\,^\circ$ und zwei Störquellen: eine bei -20 ° und eine bei 60 °.

krophonrauschen, die diffuse Störung bei höheren Frequenzen und auch bei gerichteten Störquellen für höhere Nachhallzeiten (vgl. Bild 6.11).

- Unterschiedlich gewähltes SNR bei gerichteten Störschallquellen hat insofern Auswirkungen, da hier das Verhältnis der Störleistung des räumlich korrelierten zum räumlich unkorrelierten Rauschen maßgeblich ist. Je größer dieses Verhältnis ausfällt, desto größer ist auch die erzielbare Störgeräuschunterdrückung (vgl. Bild 6.10).

- Weiterhin gilt für alle GSC-Strukturen, dass mit längeren Filterimpulsantworten in der *Adaptive Noise Cancellation* für höhere Nachhallzeiten auch eine höhere Störgeräuschunterdrückung erreichbar ist.

- Interessant erscheint hier noch eine explizite Untersuchung der Anzahl der verwendeten Filterkoeffizienten in den adaptiven *Blocking-Matrix*-Realisierungen, für die im Folgenden einige exemplarische Ergebnisse präsentiert werden sollen.

Für das *Szenario-2* wurden unterschiedliche Werte $B \in \{64, 128, 256, 512\}$ für die Anzahl der Filterkoeffizienten bei einer Nachhallzeit von $T_{60} = 0{,}3\,\mathrm{s}$ gewählt. In Bild 8.15 ist zunächst die *Blocking Ability* in (a) und der SNR-Gewinn in (b) dargestellt. Für die Verfahren mit TFRBM, GTTRBM und GEVBM sind geringe Unterschiede für unterschiedliche Werte B zu erkennen. Insgesamt scheint tendenziell eine größer gewählte Filterlänge zu einer höheren Sprachsignalunterdrückung der *Blocking-Matrix*-Strukturen zu führen und zu einem schlechteren SNR-Gewinn des entsprechenden GSCs. Auffällig sind die Ergebnisse für das Referenzsystem mit ASCBM. Hier scheint sich die eher geringe Frequenzauflösung bei $B = 64$ stärker auszuwirken und führt zu leicht schlechteren Ergebnissen im Vergleich zur GEVBM. Generell kann noch angemerkt werden, dass wenn durch den *Fixed Beamformer* der direkte Pfad nicht kohärent aufsummiert wird, umso mehr Filterkoeffizienten in der ASCBM notwendig werden, um die gleiche Sprachsignaldämpfung zu erzielen. Da die TFRBM, GTFRBM und GEVBM unabhängig vom FBF arbeiten, ist hierin ein klarer Vorteil zu sehen. Dass $B = 64$ für die ASCBM eher ungünstig scheint, wird auch durch die spektrale Varianz σ^2_{LDS} in Bild 8.16 (a) bestätigt. Denn für diese kurze Filterlänge ergeben sich die größten spektralen Abweichungen zwischen dem GSC- und dem FBF-Ausgangssignal. Ab $B \geq 128$ stellen sich jedoch wieder

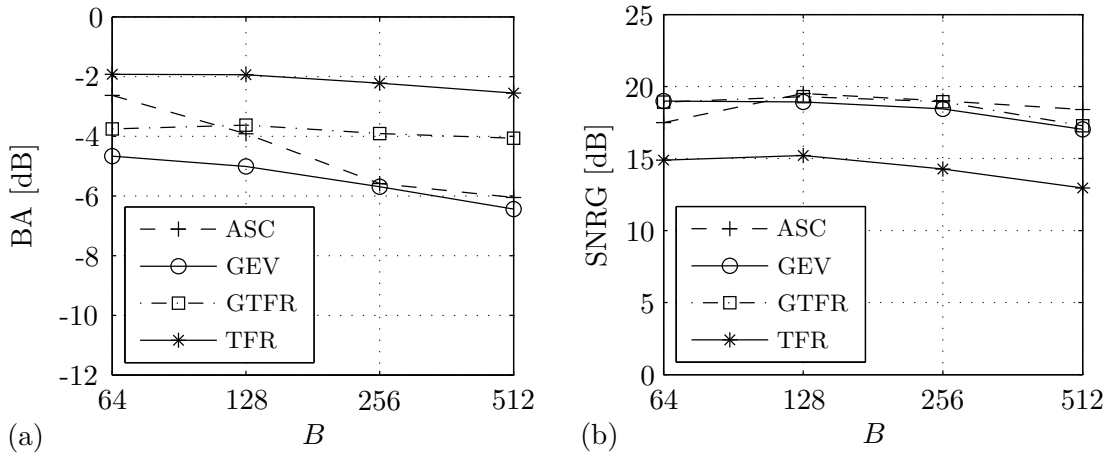

Bild 8.15: *Blocking Ability* in (a) und SNR-Gewinn in (b) für eine Sprecherrichtung von $\theta_s = 45\,^\circ$ und eine Störquelle bei $\theta_n = -20\,^\circ$ für unterschiedliche Filterlängen der *Blocking Matrix*.

die geringsten Werte für die Varianz im Vergleich zu den anderen Verfahren ein. Daher sollte der GSC mit ASCBM als Referenzsystem für die Messung der Sprachqualität mit $B = 64$ als fragwürdig gelten. Denn auch die PSM-Werte in Bild 8.16 (b) liegen für die eigenvektorbasierten Verfahren für $B = 64$ unter denen bei $B = 128$. Insgesamt wird hier in Übereinstimmung mit [HK02] eine Filterlänge von 128 oder 256 als sinnvoll erachtet.

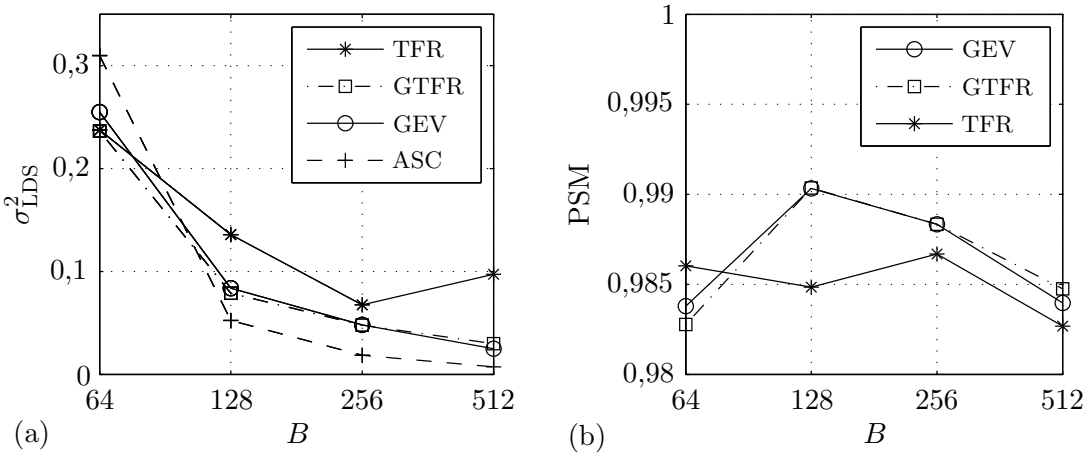

Bild 8.16: In (a) Varianz der Verhältnisse der spektralen Leistungsdichte des GSC-Ausgangssignals zu dem FBF Ausgangssignal und in (b) das perzeptive Qualitätsmaß für eine Sprecherrichtung von $\theta_s = 45\,^\circ$ und einer Störquelle bei $\theta_n = -20\,^\circ$ für unterschiedliche Filterlängen der *Blocking Matrix*.

8.4.2 Blinder Generalized Sidelobe Canceller

Den vorangegangenen Simulationsergebnissen mit einem DSB als *Fixed Beamformer* im GSC folgen nun Experimente, bei denen das *Matched Filter* Gl. (8.54) als *Fixed Beamformer* mit den eigenvektorbasierten *Blocking-Matrix*-Methoden kombiniert wird. Für diese Anordnungen ist dann keine explizite Schätzung der Sprecherrichtung mehr erforderlich. In den nachfolgenden Bildern 8.17 bis 8.20 sind für die unterschiedlichen Szenarien die SNR-Gewinne und die PSM-Werte für den DSB und GTFRBM bzw. GEVBM sowie für den *Matched Filter* und

GTFRBM bzw. GEVBM dargestellt. Letztere sind gekennzeichnet durch "MF-GTFR" bzw. "MF-GEV". Dabei zeigen die SNR-Gewinne für die beiden Varianten des *Fixed Beamformers* und jeweils gleicher *Blocking Matrix* durchweg fast identische Verläufe. Lediglich die PSM-Werte liefern für größere Nachhallzeiten leichte Differenzen zu Ungunsten der "blinden" Varianten mit *Matched Filter Beamformer* auf. Diese ergeben sich durch eine minimale Anhebung der oberen Frequenzkomponenten, welche bei subjektiven Hörtests aber nicht als störend empfunden wurde.

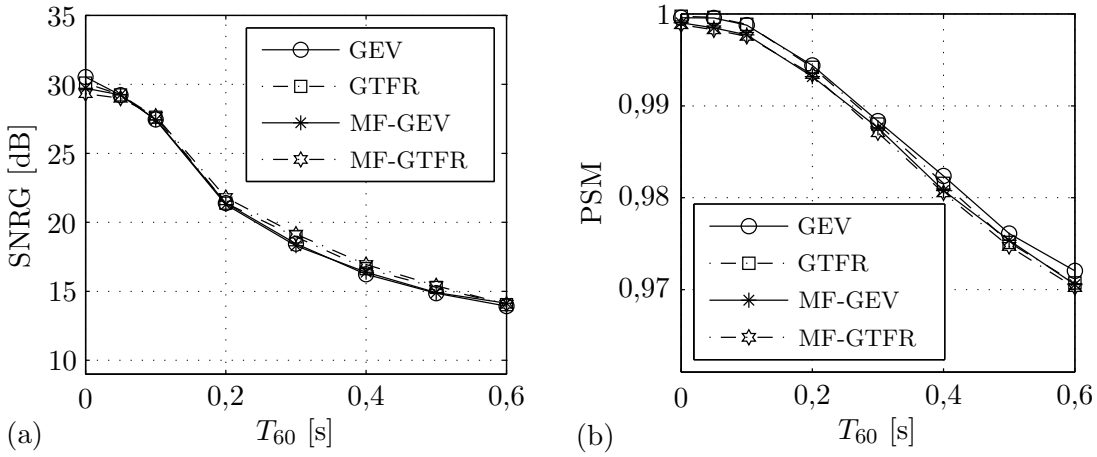

Bild 8.17: Vergleich zwischen DSB und *Matched Filter* als *Fixed Beamformer*: SNR-Gewinn in (a) und das perzeptive Qualitätsmaß in (b) für eine Sprecherrichtung von $\theta_s = 45\,^\circ$ und einer Störquelle bei $\theta_n = -20\,^\circ$.

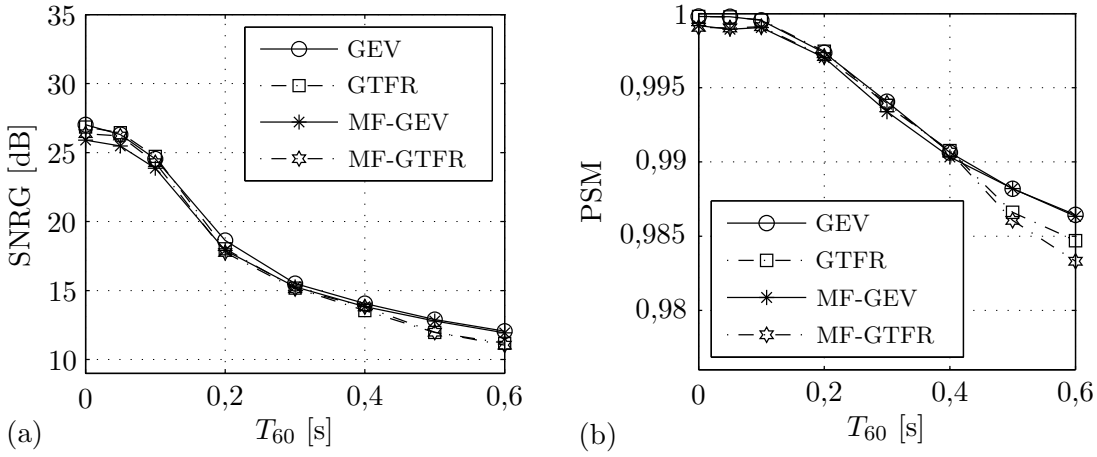

Bild 8.18: Vergleich zwischen DSB und *Matched Filter* als *Fixed Beamformer*: SNR-Gewinn in (a) und das perzeptive Qualitätsmaß in (b) für eine Sprecherrichtung von $\theta_s = 0\,^\circ$ und eine Störquelle bei $\theta_n = 60\,^\circ$.

Die guten Ergebnisse in den Bildern 8.17 bis 8.20 des blinden *Generalized Sidelobe Cancellers* im Vergleich zu der Variante mit DSB als *Fixed Beamformer* und damit implizit der Vergleich zur klassichen Variante nach Griffiths und Jim [GJ82] bestätigen exemplarisch dessen Leistungsfähigkeit. Insbesondere, da die DORBM und der DSB als optimal angesetzt wurden. In der Regel können beim Schätzen der Sprecherrichtung jedoch Fehler auftreten, wodurch die DORBM und der DSB keine optimalen Signale liefern. Dies soll abschließend

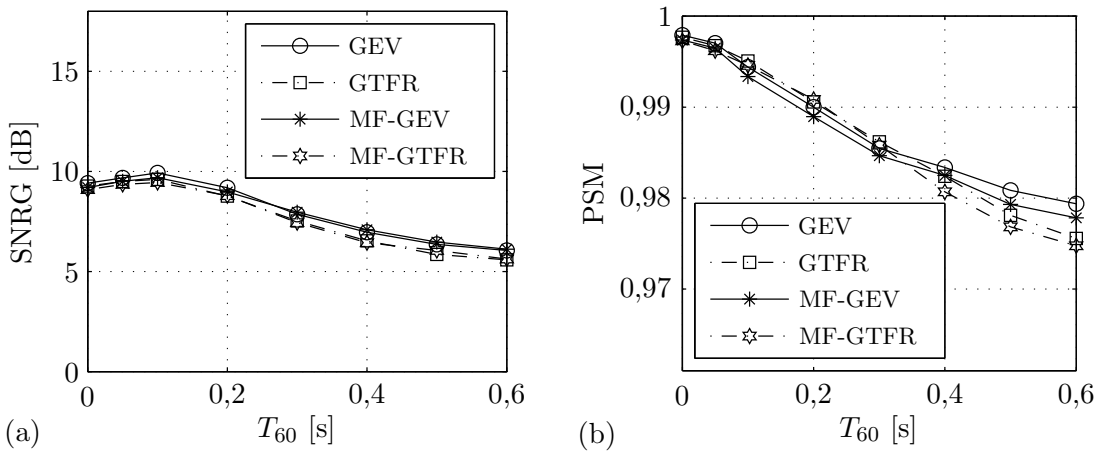

Bild 8.19: Vergleich zwischen DSB und *Matched Filter* als *Fixed Beamformer*: SNR-Gewinn in (a) und das perzeptive Qualitätsmaß in (b) für eine Sprecherrichtung von $\theta_s = 0\,^\circ$ und zwei Störquellen: eine bei -20 ° und eine bei 60 °.

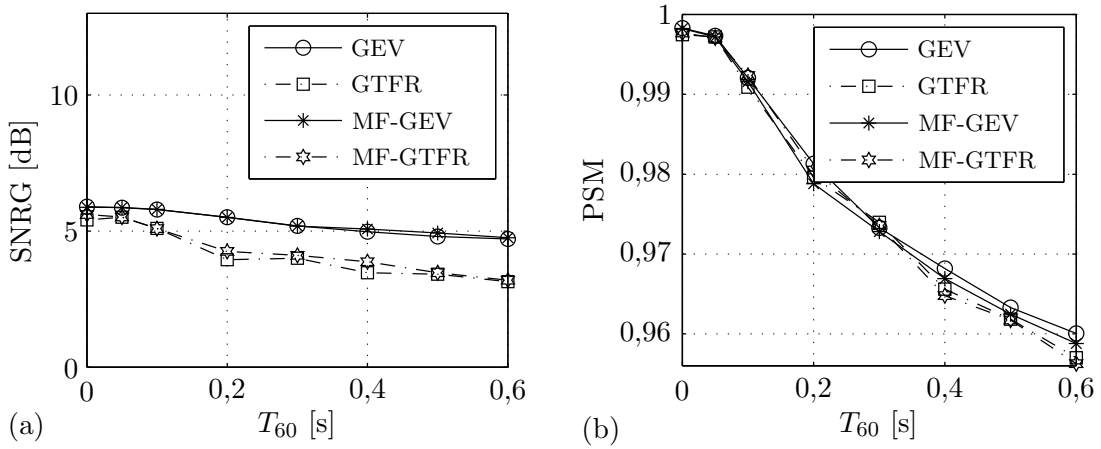

Bild 8.20: Vergleich zwischen DSB und *Matched Filter* als *Fixed Beamformer*: SNR-Gewinn in (a) und das perzeptive Qualitätsmaß in (b) für eine Sprecherrichtung von $\theta_s = 45\,^\circ$ und einem diffusen Störschallfeld.

für das *Szenario-3* mit einer Broadside-Ausrichtung $\theta_t = 0°$ des DSBs für variierende geringe Abweichungen $\Delta\theta \in \{5°, 10°, 15°\}$ von der tatsächlichen Sprecherrichtung

$$\theta_t = \theta_s + \Delta\theta \tag{8.61}$$

gezeigt werden. Um die Ergebnisse in etwas kompakterer Form darzustellen soll lediglich die Abweichung zwischen GSC mit DORBM und dem blinden GSC mit *Matched Filter* und GEVBM präsentiert werden. Der in Bild 8.21 (a) gezeigte Unterschied der *Blocking Ability* ergibt sich im logarithmischen Maßstab zu

$$\Delta\mathrm{BA} = (\mathrm{BA}|_{\text{MF-GEV}} - \mathrm{BA}|_{\text{DORBM}})\,\mathrm{dB} \tag{8.62}$$

und die in Bild 8.21 (b) dargestellte Differenz der SNR-Gewinne ist folglich

$$\Delta\mathrm{SNRG} = (\mathrm{SNRG}|_{\text{MF-GEV}} - \mathrm{SNRG}|_{\text{DORBM}})\,\mathrm{dB}. \tag{8.63}$$

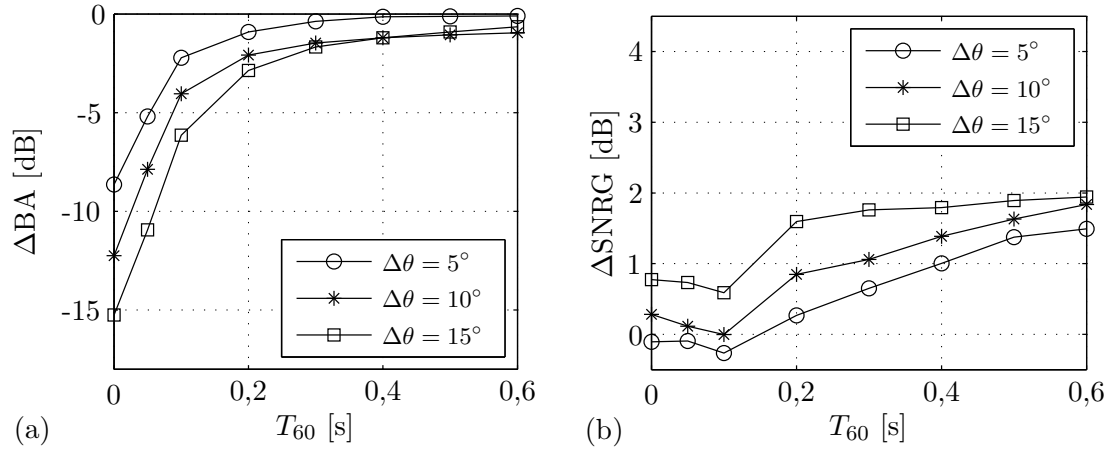

Bild 8.21: Differenzen der *Blocking Ability* in (a) und des SNR-Gewinns in (b) zwischen GSC mit DSB und DORBM und GSC mit MF und GEVBM für unterschiedliche Sprecherrichtungen mit den Abweichungen 5 °, 10 ° und 15 ° relativ zu Broadside. Der DSB ist jeweils auf $\theta_t = 0\,°$ eingestellt. Die Störquelle befindet sich bei $\theta_n = 60\,°$.

An den Verläufen in Bild 8.21 (a) ist gut zu erkennen, dass mit einer größer werdenden Abweichung $\Delta\theta = \theta_t - \theta_s$ die DORBM eine geringer werdende Dämpfungseigenschaft bezüglich des Sprachsignals besitzt. Andersherum kann gesagt werden, dass die Dämpfungseigenschaft der GEVBM nahezu gleich bei relativ geringer Variation der Sprecherrichtung ist. Aus diesen Zusammenhängen heraus sind dann die Verläufe des SNR-Gewinns in Bild 8.21 (b) folgerichtig. Denn mit steigendem $\Delta\theta$ nimmt das relative SNR des blinden GSCs im Vergleich zur konventionellen Methode zu.

8.5 Zusammenfassung

In diesem Kapitel wurde die Struktur des *Generalized Sidelobe Cancellers*[4] bestehend aus einem *Fixed Beamformer* zur Erzeugung eines Sprachreferenzsignals, einer *Blocking Matrix* zur Erzeugung eines Rauschreferenzsignals und einer *Adaptive Noise Cancellation* zur Minimierung des Rauschens im Ausgang des *Fixed Beamformers* basierend auf dem Rauschreferenzsignal erläutert. Ausgehend von den vorangegangenen Erkenntnissen zum statistisch optimalen *Beamforming* mittels adaptiver Eigenwertzerlegung im Frequenzbereich wurden hier zwei neue Methoden zur Bildung einer *Blocking Matrix* vorgestellt. Zum Einen ist dies die GTFRBM, welche aus Verhältnissen von geschätzten Raumübertragungsfunkionen besteht und äquivalent zur BM nach Gannot et al. [GBW01] ist. Jedoch erfolgt bei dem hier vorgeschlagenen Verfahren die Schätzung der Verhältnisse der Raumübertragungsfunktionen im Gegensatz zu [GBW01] mit Hilfe einer Eigenwertzerlegung. Die zweite neuartige Methode GEVBM wird ebenfalls mit Hilfe des dominanten Eigenvektors bestimmt, jedoch basierend auf dem Orthogonalitätsprinzip in Anlehnung an das Verfahren nach Hoshuyama et al. [HSH99]. Beide Matrizen, GTFRBM und GEVBM, weisen in Kombination mit dem DSB und

[4]Eine GSC-Implementierung in C/C++ bestehend aus einer Sprecherrichtungsbestimmung mit Hilfe des dominanten Eigenvektors, einem DSB als *Fixed Beamformer*, der GEV *Blocking Matrix* und dem ANC für fünf Mikrophone und den zuvor angegebenen Filterlängen weist für die Rechenzeit einen Echtzeitfaktor von ca. 0,3 mit einem Intel Quad-Core Xeon E5345/2,33 GHz Prozessor auf. Hierin ist das mehrkanalige Ein- und Ausgabemanagement bereits enthalten.

der ANC eine bessere Störgeräuschreduktion im Vergleich zu dem Verfahren nach Gannot et al. [GBW01] und der konventionellen Methode nach Griffiths und Jim [GJ82] auf. Im Allgemeinen liefert die Variante GEVBM ein geringfügig besseres SNR und gleichzeitig weniger Sprachverzerrungen im Vergleich zur Methode mit GTFRBM. Weiterhin wurde in diesem Kapitel ein *Matched Filter Fixed Beamformer* mit den eigenvektorbasierten BM-Varianten kombiniert und die Gesamtanordnung als blinder GSC bezeichnet. Die resultierenden Vorteile sind dabei wie folgt: Zum einen kann jede eigenvektorbasierte BM auch bei gleichzeitig zum Sprecher aktivem stationären Rauschen berechnet werden. Dies ist zwar mit dem Verfahren nach Gannot et al. [GBW01] auch möglich, die hier vorgeschlagenen Methoden führen jedoch zu einer höheren Störgeräuschunterdrückung und weniger Sprachverzerrungen. Und zum anderen wird beim *Matched Filter* FBF keine explizite Sprecherrichtungsbestimmung benötigt, da dieser auf den adaptiv berechneten dominanten Eigenvektoren basiert. Es ergibt sich dabei zwar eine leicht größere Sprachverzerung als bei der Variante mit einem perfekten DSB, aber es entsteht der Vorteil einer Reduzierung des Rechenaufwandes.

Kapitel 9

Zusammenfassung

Im Rahmen dieser Arbeit wurden Algorithmen zur mehrkanaligen Störgeräuschreduktion basierend auf der Lösung eines Eigenwertproblems im Frequenzbereich entwickelt und untersucht. Das betrachtete Eigenwertproblem entsteht aufgrund eines Optimierungsproblems, welchem die Maximierung des Signal-zu-Rauschleistungsverhältnisses am *Beamformer*-Ausgang zugrunde liegt. Die Lösung des Eigenwertproblems kam hierbei in zwei *Beamformer*-Strukturen zum Tragen: zum einen als *Filter-and-Sum-Beamformer* und zum anderen als *Generalized Sidelobe Canceller*, bestehend aus den Komponenten *Fixed Beamformer*, *Blocking Matrix* und Adaptive Sidelobe Canceller, wobei der neuartige Ansatz in der *Blocking Matrix* und im *Fixed Beamformer* angesetzt wurde. Grundsätzlich ermöglicht der *Generalized Sidelobe Canceller* eine höhere Störgeräuschreduktion als der *Filter-and-Sum-Beamformer*, setzt jedoch im Vergleich zu diesem eine gewisse Stationarität der Sprecherposition voraus.

In einem adaptiven *Filter-and-Sum-Beamformer* zur breitbandigen Sprachsignalverbesserung kam das Kriterium der Maximierung des Signal-zu-Rauschleistungsverhältnisses aufgrund der einhergehenden Signalverzerrungen bislang nicht zum Einsatz. In der vorliegenden Arbeit ist es gelungen, durch geeignete Nachfilterverfahren die entstehenden Sprachverzerrungen deutlich zu reduzieren und somit eine Anwendung zur mehrkanaligen Störgeräuschreduktion zu ermöglichen. Basierend auf diesen Verfahren ist ein *Matched Filter* als Teil eines neuartigen *Generalized Sidelobe Cancellers* entstanden. Dieser beinhaltet desweiteren eine eigenentwickelte *Blocking Matrix*, welcher ebenfalls das Kriterium der Maximierung des Signal-zu-Rauschleistungsverhältnisses zugrunde liegt.Signal-zu-

Die in dieser Arbeit vorgelegten *Beamforming*-Verfahren, sowohl *Filter-and-Sum-Beamformer* als auch *Generalized Sidelobe Canceller*, zeichnen sich insbesondere durch ihre blinden Adaptionseigenschaften aus. Dies bedeutet, dass keine explizite Positionsbestimmung des Sprechers notwendig ist und die geometrische Anordnung der Mikrophone unbekannt sein kann. Weiterhin erfolgt bei der Adaption eine implizite, konstruktive Nutzung mehrerer Ausbreitungspfade des Sprachsignals zwischen dem Sprecher und der Mikrophongruppe.

Der Vergleich unterschiedlicher Ansätze zum statistisch optimalen *Beamforming* in Kapitel 4 zeigte, dass die Lösungen sich nur in einem skalaren Faktor unterscheiden. Daraus entstand der grundlegende Gedanke zur Realisierung eines *Filter-and-Sum-Beamformers* mittels SNR-Maximierung und einer nachgeschalteten Normalisierung der resultierenden Filterkoeffizienten. Ziel der Normalisierung war es, eine approximative Darstellung eines MVDR *Beam-*

formers zu erreichen. Im Gegensatz zu dem MVDR *Beamformer* bietet der neue Ansatz jedoch den Vorteil auf eine Positionsbestimmung des Sprechers zu verzichten. Ein weiterer Vorteil der Lösung des Eigenwertproblems ist die Einbeziehung der Halleigenschaften von Räumen, wie in den Simulationen in Kapitel 4 gezeigt werden konnte. Ausgehend von der Analyse der Kohärenz unterschiedlicher Störgeräuschfelder in Kapitel 2 wurde in Kapitel 5 aufgezeigt, wie die Formulierung des zu lösenden Eigenwertproblems ausfällt: Für den Fall von räumlich korrelierten Störungen wie diffuse und gerichtete Störschallfelder ergibt sich das verallgemeinerte Eigenwertproblem bezüglich der Kreuzleistungsdichtematrix der Störsignale und der Kreuzleistungsdichtematrix aus der Überlagerung von Störsignal- und Sprachsignalkomponenten. Bei räumlich unkorrelierten Störungen wie Mikrophonrauschen folgt hingegen das spezielle Eigenwertproblem bezüglich der Matrix der Kreuzleistungsdichten der Sprachsignale an den Mikrophonen. Da jedoch bei einem diffusen Störschallfeld in Abhängigkeit von der Mikrophonanordnung eine signifikante Kohärenz primär für den unteren Frequenzbereich vorliegt, wird für dieses Störschallfeld ebenfalls die Lösung des speziellen Eigenwertproblems empfohlen. Dadurch fällt der zu erwartende SNR-Gewinn im unteren Frequenzbereich zwar geringer aus, jedoch ergibt sich der Vorteil einer reduzierten Rechenkomplexität. Zur Bestimmung der jeweiligen spektralen Kreuzleistungsdichtematrizen ist eine robuste Sprache/Pause-Detektion notwendig. Ein geeignetes Verfahren hierzu wurde im Anhang in Kapitel D vorgestellt.

In Kapitel 5 wurden iterative Verfahren zur Bestimmung eines Eigenvektors korrespondierend zum größten Eigenwert eines speziellen und des verallgemeinerten Eigenwertproblems präsentiert und miteinander verglichen. Dies waren zum einen eigenentwickelte Gradientenverfahren und zum anderen Verfahren aus der Literatur, sowohl Gradienten- als auch Fixpunktverfahren. Hierbei zeigten die experimentellen Ergebnisse eine deutliche Überlegenheit der Fixpunktverfahren im Vergleich zu den Gradientenverfahren für die Problemstellung des verallgemeinerten Eigenwertproblems. Für das spezielle Eigenwertproblem zeigt das neuartige Gradientenverfahren einerseits eine signifikante Robustheitssteigerung bezüglich der Konvergenz im Vergleich zu dem Gradientenverfahren nach Oja und andererseits ähnlich gute Konvergenzeigenschaften wie die Fixpunktverfahren auf, mit dem Vorteil einer deutlichen Verringerung der Rechenkomplexität. Für das akustische *Beamforming* unter Berücksichtigung der Kreuzleistungsdichtematrix der Störung wird daher ein Fixpunktverfahren und beim Einsatz eines *Beamformers* in einer Umgebung, in der außer dem Sprecher keine weiteren dominanten Schallquellen zu erwarten sind, das eigenentwickelte Gradientenverfahren präferiert.

Bei der Anwendung des dominanten Eigenvektors zur akustischen Strahlformung als *Filter-and-Sum-Beamformer* sind in Kapitel 6 die resultierenden Sprachverzerrungen untersucht worden. Dabei kamen die in Kapitel 3 eingeführten Bewertungskriterien zum Einsatz, insbesondere die wahrnehmungsbasierte Qualitätsbewertung PEMO-Q. Die vorgestellten drei eigenentwickelten Verfahren zur Normalisierung der Filterkoeffizienten wiesen eine signifikante Reduzierung der Sprachverzerrung auf, wobei die blinde analytische Normalisierung die besten Ergebnisse für alle betrachteten akustischen Szenarien zeigte. Aufgrund der guten Adaptionseigenschaften des neuen *Beamforming*-Verfahrens und der Verwendung kurzer Filterlängen ist das Folgen einer variierenden Sprecherposition möglich.

Bei der Realisierung des neuartigen *Generalized Sidelobe Cancellers* in Kapitel 8 findet die Lösung eines Eigenwertproblems im Frequenzbereich insbesondere im Teilkomplex der *Blocking Matrix* ihre Anwendung. Hier war der Grundgedanke, äquivalent zu der *Blocking Matrix* nach Hoshuyama, einen zum Sprachsignal orthogonalen Unterraum mittels eines Sprach-

referenzsignals zu erzeugen. Im Gegensatz zu der *Blocking Matrix* nach Hoshuyama ist jedoch kein explizites Sprachreferenzsignal erforderlich, da dies inhärenter Bestandteil des neuen Algorithmus ist. Die neuartige *Blocking Matrix* bietet somit den Vorteil, dass keine Sprecherrichtungsbestimmung notwendig ist und eine Adaption auch in stark gestörten Umgebungen mit permanent aktiven Störschallquellen möglich ist. Diese Vorzüge bietet die *Blocking Matrix* nach Gannot zwar auch, jedoch weist diese deutliche Sprachverzerrungen und eine geringere Störgeräuschreduktion im Vergleich zu der Eigenentwicklung auf. Die klassische Variante der *Blocking Matrix* nach Griffiths und Jim kann zwar ebenfalls bei permanentem Störschallfeld betrieben werden, hat jedoch zur eigenentwickelten Methode den Nachteil, dass nur der direkte Ausbreitungspfad des Sprachsignals berücksichtigt wird.

Der in der GSC-Struktur notwendige *Fixed Beamformer* wurde in zwei Varianten umgesetzt: zum einen als *Delay-and-Sum-Beamformer* und zum anderen mittels eines eigenentwickelten *Matched Filters*. Die für den DSB erforderliche Sprecherrichtung wurde mit einem neuartigen Verfahren, ebenfalls basierend auf dem dominanten Eigenvektor, ermittelt. Dieses Verfahren zeigt im Gegensatz zu den in der Literatur diskutierten Methoden den Vorteil, nahezu unabhängig von dem betrachteten Störgeräuschfeld zu sein, wie die experimentellen Ergebnisse in Kapitel 7 demonstrieren. Das *Matched Filter* als *Fixed Beamformer* weist zwar im Gegensatz zum DSB leichte Sprachverzerrungen auf, bietet jedoch den Vorteil einen blinden *Generalized Sidelobe Canceller* zu realisieren: es ist keine Sprecherrichtungsbestimmung notwendig und die geometrische Anordnung der Mikrophone kann unbekannt sein.

Anhang A

Lineare Algebra – Matrizen

Im Folgenden sollen einige grundlegende Begriffe bezüglich der in dieser Arbeit verwendeten Matrix-Algebra definiert werden.

A.1 Grundlagen

Rang Für eine Matrix $\mathbf{A}$ der Dimension $(m \times n)$ stimmt die maximale Anzahl linear unabhängiger Spalten (Spaltenrang) mit der maximalen Anzahl linear unabhängiger Zeilen (Zeilenrang) überein und wird kurz als Rang bezeichnet

$$\text{Rang}(\mathbf{A}) \leq \min\{m, n\}. \tag{A.1}$$

Spur Die Summe über alle Hauptdiagonalenelemente a_{ii} mit $i = 1, 2, \ldots, m$ einer Matrix $\mathbf{A}$ der Dimension $(m \times m)$ wird Spur genannt

$$\text{Spur}(\mathbf{A}) = \sum_{i=1}^{m} a_{ii}. \tag{A.2}$$

Hermitesch Eine komplexe, quadratische Matrix $\mathbf{A}$ heißt hermitesch, wenn sie gleich der konjugierten, transponierten Matrix $\mathbf{A}$ ist

$$\mathbf{A} = (\mathbf{A}^*)^T = \mathbf{A}^H. \tag{A.3}$$

Unitär/Orthogonal Dies ist eine Bezeichnung für eine komplexwertige, quadratische Matrix $\mathbf{A}$, wenn deren Spalten zueinander orthonormal sind. Damit gilt

$$\mathbf{A}^H\mathbf{A} = \mathbf{I}, \tag{A.4}$$

mit $\mathbf{I}$ für die Einheitsmatrix und weiterhin für die Inverse

$$\mathbf{A}^{-1} = \mathbf{A}^H. \tag{A.5}$$

Ist $\mathbf{A}$ eine reelwertige Matrix, die die Eigenschaften Gl. (A.4) und Gl. (A.5) erfüllt, so wird sie als orthogonal bezeichnet.

Kern/Bild Gegeben sei die lineare Abbildung $\mathbf{A} : \mathbf{V} \rightarrow \mathbf{W}$. Für den Kern der Abbildung gilt

$$\text{Kern}(\mathbf{A}) = \{\mathbf{v} \in \mathbf{V} : \mathbf{0} = \mathbf{A}\mathbf{v}\} \tag{A.6}$$

und die Menge der Vektoren aus $\mathbf{W}$, die die Abbildung tatsächlich annimmt, wird Bild genannt

$$\text{Bild}(\mathbf{A}) = \{\mathbf{w} \in \mathbf{W} : \mathbf{w} = \mathbf{A}\mathbf{v}, \mathbf{v} \in \mathbf{V}\}. \tag{A.7}$$

Ableitung bezüglich eines komplexen Vektors Es sei gegeben der komplexe Vektor $\mathbf{F} = [F_1, F_2, \ldots, F_m]^T$ der Dimension $(m \times 1)$. Die Elemente des Vektors bestehen aus $F_i = x_i + j \cdot y_i$, $i = 1, 2, \ldots, m$ mit den reellwertigen Größen x_i und y_i und der imaginären Einheit j. Dann ist $\partial/\partial\mathbf{F}$ die Ableitung bezüglich $\mathbf{F}$ und $\partial/\partial\mathbf{F}^*$ die korrespondierende komplexe konjugierte Ableitung

$$\frac{\partial}{\partial \mathbf{F}} = \frac{1}{2}\begin{bmatrix} \frac{\partial}{\partial x_1} - \frac{\partial}{\partial y_1} \\ \frac{\partial}{\partial x_2} - \frac{\partial}{\partial y_2} \\ \vdots \\ \frac{\partial}{\partial x_m} - \frac{\partial}{\partial y_m} \end{bmatrix} \quad \frac{\partial}{\partial \mathbf{F}^*} = \frac{1}{2}\begin{bmatrix} \frac{\partial}{\partial x_1} + \frac{\partial}{\partial y_1} \\ \frac{\partial}{\partial x_2} + \frac{\partial}{\partial y_2} \\ \vdots \\ \frac{\partial}{\partial x_m} + \frac{\partial}{\partial y_m}. \end{bmatrix} \tag{A.8}$$

Mit Hilfe von Gl. (A.8) kann folgender Gradientenvektor definiert werden

$$\nabla_{\mathbf{F}^*} = 2\frac{\partial}{\partial \mathbf{F}^*}. \tag{A.9}$$

Span Die lineare Hülle (auch engl. linear span) bildet einen Vektorraum aus einer vorgegebenen Menge von Vektoren $\{\mathbf{v}_i : i = 1, \ldots, m\}$ durch deren Linearkombinationen

$$\text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m) = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_m\mathbf{v}_m : a_1, a_2, \ldots, a_m \in \mathbb{C}\}. \tag{A.10}$$

Matrix Inversion Lemma Es seien $\mathbf{A}$ und $\mathbf{B}$ zwei positiv definite $(M \times M)$-Matrizen, $\mathbf{D}$ sei positiv definit der Dimension $(N \times N)$ und $\mathbf{C}$ ist eine $(M \times N)$-Matrix. Dann gilt für

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^H \tag{A.11}$$

das Matrix Inversion Lemma[1] [Hay02]

$$\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}[\mathbf{D} + \mathbf{C}^H\mathbf{B}\mathbf{C}]^{-1}\mathbf{C}^H\mathbf{B}. \tag{A.12}$$

A.2 Matrix Inversion für optimales Beamforming

An dieser Stelle wird zum einen die Äqivalenz des MV-Ansatzes nach Gl. (4.21) und der Lösung Gl. (4.28) zum Lösungsansatz

$$\underset{\mathbf{F}(\Omega)}{\text{minimiere}} \qquad \mathbf{F}^H(\Omega)\mathbf{\Phi}_{\mathbf{XX}}(\Omega)\mathbf{F}(\Omega) \tag{A.13}$$

$$\text{mit} \qquad \mathbf{F}^H(\Omega)\mathbf{H}(\Omega) = 1 \tag{A.14}$$

[1] Das Matrix Inversion Lemma ist in der Literatur ebenfalls unter Sherman-Morrison-Woodbury oder Woodbury Formel bzw. Woodbury Matrix Identität bekannt.

mit den resultierenden Filterkoeffizienten

$$\mathbf{F}_{\text{Frost}}(\Omega) = \frac{\mathbf{\Phi}_{\mathbf{XX}}^{-1}(\Omega)\mathbf{H}(\Omega)}{\mathbf{H}^H(\Omega)\mathbf{\Phi}_{\mathbf{XX}}^{-1}(\Omega)\mathbf{H}(\Omega)}. \tag{A.15}$$

gezeigt, welche nach Frost [Fro72] mit Hilfe eines Gradienten-Abstiegs-Verfahrens berechnet werden können. Zum Anderen wird die faktorisierte MMSE-Lösung Gl. (4.45) hergeleitet. Grundlage in beiden Fällen ist die Invertierung der Matrix $\mathbf{\Phi}_{\mathbf{XX}}(\Omega)$, wobei im Folgenden auf die frequenzabhängige Notation – gekennzeichnet durch den Parameter Ω – verzichtet werden soll. Zur Invertierung von

$$\mathbf{\Phi}_{\mathbf{XX}} = \phi_{S_cS_c}\mathbf{H}\mathbf{H}^H + \mathbf{\Phi}_{\mathbf{NN}} \tag{A.16}$$

sind die Matrizen in Gl. (A.11) zu definieren als:

$$\mathbf{B}^{-1} = \mathbf{\Phi}_{\mathbf{NN}}, \quad \mathbf{C} = \sqrt{\phi_{S_cS_c}}\mathbf{H}, \quad \mathbf{D} = 1. \tag{A.17}$$

Die Anwendung von Gl. (A.12) auf Gl. (A.16) ergibt

$$\left[\phi_{S_cS_c}\mathbf{H}\mathbf{H}^H + \mathbf{\Phi}_{\mathbf{NN}}\right]^{-1} = \mathbf{\Phi}_{\mathbf{NN}}^{-1} - \frac{\phi_{S_cS_c}\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}\mathbf{H}^H\mathbf{\Phi}_{\mathbf{NN}}^{-1}}{1 + \phi_{S_cS_c}\mathbf{H}^H\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}}. \tag{A.18}$$

Mit

$$\left[\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}\mathbf{H}^H\mathbf{\Phi}_{\mathbf{NN}}^{-1}\right]\mathbf{H} = \left[\mathbf{H}^H\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}\mathbf{\Phi}_{\mathbf{NN}}^{-1}\right]\mathbf{H} \tag{A.19}$$

folgt weiter

$$\left[\phi_{S_cS_c}\mathbf{H}\mathbf{H}^H + \mathbf{\Phi}_{\mathbf{NN}}\right]^{-1}\mathbf{H} = \left[\mathbf{\Phi}_{\mathbf{NN}}^{-1} - \frac{\phi_{S_cS_c}\mathbf{H}^H\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}\mathbf{\Phi}_{\mathbf{NN}}^{-1}}{1 + \phi_{S_cS_c}\mathbf{H}^H\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}}\right]\mathbf{H} \tag{A.20}$$

$$= \frac{1}{1 + \phi_{S_cS_c}\mathbf{H}^H\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}}\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}. \tag{A.21}$$

Lösung nach Frost [Fro72] Wird das Ergebnis der Invertierung Gl. (A.21) in Gl. (A.15) eingesetzt, so erhält man nach dem Kürzen des skalaren Faktors $1/(1 + \phi_{S_cS_c}\mathbf{H}^H\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H})$

$$\mathbf{F}_{\text{Frost}} = \frac{\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}}{\mathbf{H}^H\mathbf{\Phi}_{\mathbf{NN}}^{-1}(\Omega)\mathbf{H}} = \mathbf{F}_{\text{GMVDR}}. \tag{A.22}$$

Faktorisieren der MMSE-Lösung Um die Faktorisierung des mehrkanaligen Wiener Filters in Abschnitt 4.4 durchzuführen wird in

$$\mathbf{F}_{\text{GMMSE}} = \mathbf{\Phi}_{\mathbf{XX}}^{-1}\phi_{S_cS_c}\mathbf{H} = \left[\phi_{S_cS_c}\mathbf{H}\mathbf{H}^H + \mathbf{\Phi}_{\mathbf{NN}}\right]^{-1}\phi_{S_cS_c}\mathbf{H} \tag{A.23}$$

Gl. (A.21) eingesetzt

$$\mathbf{F}_{\text{GMMSE}} = \frac{\phi_{S_cS_c}}{1 + \phi_{S_cS_c}\mathbf{H}^H\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}}\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H} \tag{A.24}$$

$$= \left[\frac{\phi_{S_cS_c}}{\phi_{S_cS_c} + \left(\mathbf{H}^H\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}\right)^{-1}}\right]\frac{\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}}{\mathbf{H}^H\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}} \tag{A.25}$$

$$= \left[\frac{\phi_{S_cS_c}}{\phi_{S_cS_c} + \left(\mathbf{H}^H\mathbf{\Phi}_{\mathbf{NN}}^{-1}\mathbf{H}\right)^{-1}}\right]\mathbf{F}_{\text{GMVDR}}. \tag{A.26}$$

A.3 Matrix Inversion für Fixpunkt-Adaption

Das Ziel ist hier, die iterative Schätzung

$$\hat{\mathbf{\Phi}}_{\mathbf{NN},\kappa+1} = \alpha\hat{\mathbf{\Phi}}_{\mathbf{NN},\kappa} + (1-\alpha)\mathbf{N}_\kappa\mathbf{N}_\kappa^H \tag{A.27}$$

zu invertieren, wobei auf die frequenzabhängige Notation verzichtet wird, mit κ der Iterationsindex und mit α die Glättungskonstante bezeichnet ist. Die Matrizen in Gl. (A.11) werden wie folgt substituiert:

$$\mathbf{B}^{-1} = \alpha\hat{\mathbf{\Phi}}_{\mathbf{NN},\kappa}, \quad \mathbf{C} = \sqrt{1-\alpha}\mathbf{N}_\kappa, \quad \mathbf{D} = 1. \tag{A.28}$$

Nach Einsetzen der Matrizen Gl. (A.28) in Gl. (A.12) ergibt sich für

$$\hat{\mathbf{\Phi}}_{\mathbf{NN},\kappa+1}^{-1} = \frac{1}{\alpha}\left[\mathbf{I} - \frac{\hat{\mathbf{\Phi}}_{\mathbf{NN},\kappa}^{-1}\mathbf{N}_\kappa\mathbf{N}_\kappa^H}{\frac{\alpha}{1-\alpha} + \mathbf{N}_\kappa^H\hat{\mathbf{\Phi}}_{\mathbf{NN},\kappa}^{-1}\mathbf{N}_\kappa^H}\right]\hat{\mathbf{\Phi}}_{\mathbf{NN},\kappa}^{-1}. \tag{A.29}$$

Anhang B

Räumliche Kohärenz eines diffusen Schallfeldes

Ausschlaggebend für die Größe der Kohärenz ist der Phasenunterschied zwischen den Schallwellen an den Aufnahmeorten. Ist die Wellenlänge im Vergleich zum Abstand der Mikrophonsignale sehr groß, so ist der Phasenunterschied an den Empfangsorten gering und die Signale sind sich sehr ähnlich. Entsprechend der Darstellung in Bild B.1 sollen zwei Quellen in gleichem Abstand zum Mittelpunkt einer zweikanaligen Mikrophonanordnung angenommen werden, welche die beiden Signale $q_1(t)$ und $q_2(t)$ emittieren. Es soll eine Freifeldausbreitung und für die Quellen die Fernfeldnäherung gelten. Dann empfangen die beiden Sensoren die folgenden Signale

$$x_1(t) = q_1(t + \cos\varphi_1 \frac{d_{12}}{2c}) + q_2(t + \cos\varphi_2 \frac{d_{12}}{2c}) \tag{B.1}$$

$$x_2(t) = q_1(t - \cos\varphi_1 \frac{d_{12}}{2c}) + q_2(t - \cos\varphi_2 \frac{d_{12}}{2c}), \tag{B.2}$$

wobei d_{12} den Abstand zwischen den Sensoren und c die Schallgeschwindigkeit angibt. Die beiden Einfallswinkel sind beschrieben durch φ_1 bzw. φ_2. Nach der DTFT ergeben sich folglich die Signale

$$X_1(\Omega) = Q_1(\Omega)e^{j(\Omega d_{12}\cos\varphi_1)/(2Tc)} + Q_2(\Omega)e^{j(\Omega d_{12}\cos\varphi_2)/(2Tc)} \tag{B.3}$$

$$X_2(\Omega) = Q_1(\Omega)e^{-j(\Omega d_{12}\cos\varphi_1)/(2Tc)} + Q_2(\Omega)e^{-j(\Omega d_{12}\cos\varphi_2)/(2Tc)} \tag{B.4}$$

mit der normierten Kreisfrequenz Ω und der Abtastperiode T. Die komplexe Kohärenzfunktion kann äquivalend zu Gl. (2.16) angegeben werden als

$$\gamma_{X_1X_2}(\Omega) = \frac{E\{X_1(\Omega)X_2^*(\Omega)\}}{\sqrt{E\{|X_1(\Omega)|^2\}E\{|X_2(\Omega)|^2\}}}. \tag{B.5}$$

Nun soll $E\{|X_1(\Omega)|^2\} = E\{|X_2(\Omega)|^2\}$ gelten[1], so dass Gl. (B.5) mit Gl. (B.3) und Gl. (B.4) vereinfacht werden kann zu

$$\gamma_{X_1X_2}(\Omega) = \frac{1}{2}\left(e^{j\Omega d_{12}\cos\varphi_1/(Tc)} + e^{j\Omega d_{12}\cos\varphi_2/(Tc)}\right). \tag{B.6}$$

Werden also zwei Quellen mit gleicher Leistung auf einer Kugeloberflächen angeordnet, ergibt sich die Kohärenzfunktion Gl. (B.6) durch das arithmetische Mittel zweier komplexer

[1] Die Erwartungswertbildung $E\{|X_1(\Omega)|^2\}$ und $E\{|X_2(\Omega)|^2\}$ gilt bezüglich aller Realisierungen von Q_1 und Q_2.

Exponentialterme. Diese Eigenschaft kann nun auf N Quellen erweitert werden

$$\gamma_{X_1X_2}(\Omega) = \frac{1}{N}\sum_{i=1}^{N} e^{j\Omega d_{12}\cos\varphi_i/(Tc)} \tag{B.7}$$

und für unendlich viele Quellen verteilt auf einer Kugeloberfläche mit dem Radius r

$$\begin{aligned}
\gamma_{X_1X_2}(\Omega) &= \frac{1}{4\pi r^2}\int_0^{2\pi}\int_0^{\pi} e^{j\Omega d_{12}\cos\varphi/(Tc)} r^2 \sin\varphi \,\mathrm{d}\varphi\mathrm{d}\theta && \text{(B.8)}\\
&= \frac{1}{2}\int_{-1}^{1} e^{j\Omega d_{12}\vartheta/(Tc)} \,\mathrm{d}\vartheta && \text{(B.9)}\\
&= \frac{Tc}{2j\Omega d_{12}}\left(e^{j\Omega d_{12}/(Tc)} - e^{-j\Omega d_{12}/(Tc)}\right) && \text{(B.10)}\\
&= \frac{\sin(\Omega d_{12}/(Tc))}{\Omega d_{12}/(Tc)} && \text{(B.11)}\\
&= \mathrm{si}\left(\Omega\frac{d_{12}}{Tc}\right). && \text{(B.12)}
\end{aligned}$$

Das Ergebnis in Gl. (B.12) ist gerade die Kohärenzfunktion eines diffusen Schallfelds.

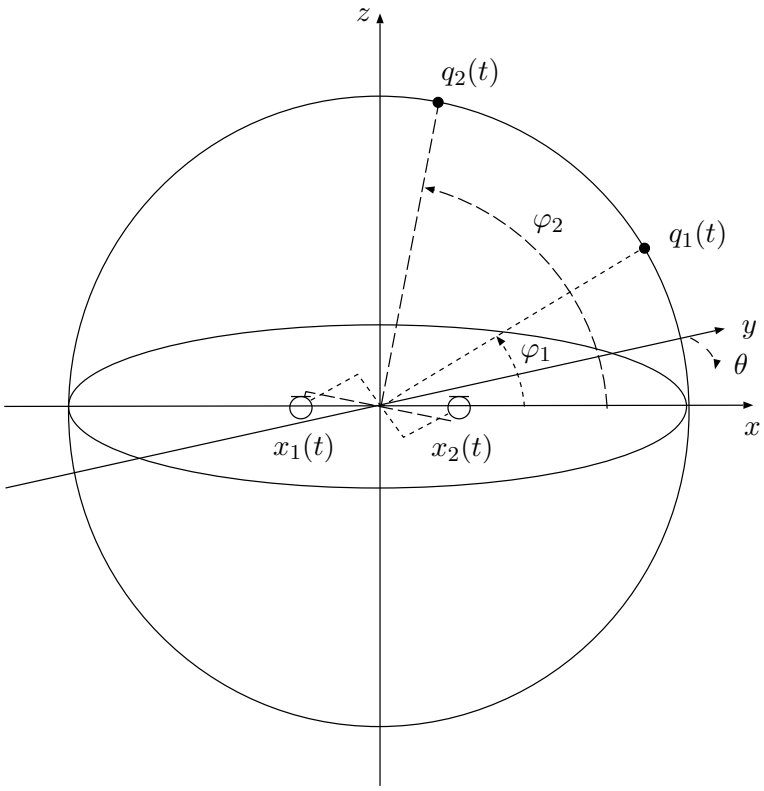

Bild B.1: Modell sphärisch angeordneter unkorrellierter Schallquellen.

Anhang C

Geometrische Anordnungen der Simulationen

In diesem Kapitel sollen die verschiedenen Simulationsumgebungen beschrieben werden, welche im Rahmen dieser Arbeit verwedet wurden. Zum einen sind dies die geometrischen Anordnungen zur Störgeräuschunterdrückung bei Anwesenheit von nur einer Sprachsignalquelle und zum anderen die geometrischen Anordnungen zur Quellentrennung bei zwei vorhandenen Sprachsignalquellen. Allgemein gilt die Abtastrate von $f_{Ab} = 12\,\text{kHz}$ für alle verwendeten Quellsignale und jeweils eine äquidistante Anordnung der Mikrophone im Abstand von $d = 4\,\text{cm}$ zueinander.

C.1 Spiegelquellenmethode für Störgeräuschunterdrückung

Zur Untersuchung der Störgeräuschunterdrückung wurden zwei Positionen für die Sprachsignalquellen gewählt, jeweils mit dem Abstand von 0,8 m zum Mittelpunkt des *Arrays*. Für die erste – gekennzeichnet durch S1 – gilt die Einfallsrichtung $\theta_{s,1} = 45°$ und die zweite – gekennzeichnet durch S2 – entsprechend $\theta_{s,2} = 0°$, jeweils relativ zu *Broadside*. Des Weiteren sind zwei Störsignalquellen jeweils im Abstand von 1,6 m zum Mittelpunkt des *Arrays* platziert, eine bei einer Richtung von $\theta_{n,1} = -20°$ – gekennzeichnet durch N1 – und die andere bei $\theta_{n,2} = 60°$ – gekennzeichnet durch N2 – ebenfalls relativ zu *Broadside*. Alle Quellen befinden sich in der gleichen Ebene auf einer Höhe von 1,5 m in einem Raum der Länge 6 m, der Breite 5 m und der Höhe 3 m. Die Anordnung in dem simulierten Raum kann dem Bild C.1 entnommen werden.

Die Signale an den Sensoren ergeben sich letztendlich durch unterschiedliche Kombinationen der Quellsignale. Grundsätzlich gilt jedoch, dass den Mischsignalen an den Mikrophonen jeweils unkorreliertes weißes Rauschen mit einem SNR von 25 dB hinzugefügt wurde. Als Nutzsignale kamen 10 Beispielsätze der TIMIT-Datenbank zum Einsatz; 5 von männlichen und 5 von weiblichen Sprechern. Die Störsignalquelle N1 bei $\theta_{n,1} = -20°$ basiert auf der Aufnahme eines PC-Lüftergeräusches und hat somit Tiefpass-Charakter. Die zweite Störquelle N2 bei $\theta_{n,2} = 60°$ ist künstlich erzeugtes weißes Rauschen mit anschließender Tiefpassfilterung. Die beiden Leistungsdichtespektren von N1 und N2 sind in Bild C.2 dargestellt. Die Kombination der verschiedenen Schallquellen ist durch folgende 4 Szenarien gegeben:

Szenario-1 Sprachquelle S1 ist aktiv (mit und ohne diffuses Störschallfeld)

Szenario-2 Sprachquelle S1 und Störquelle N1 sind aktiv

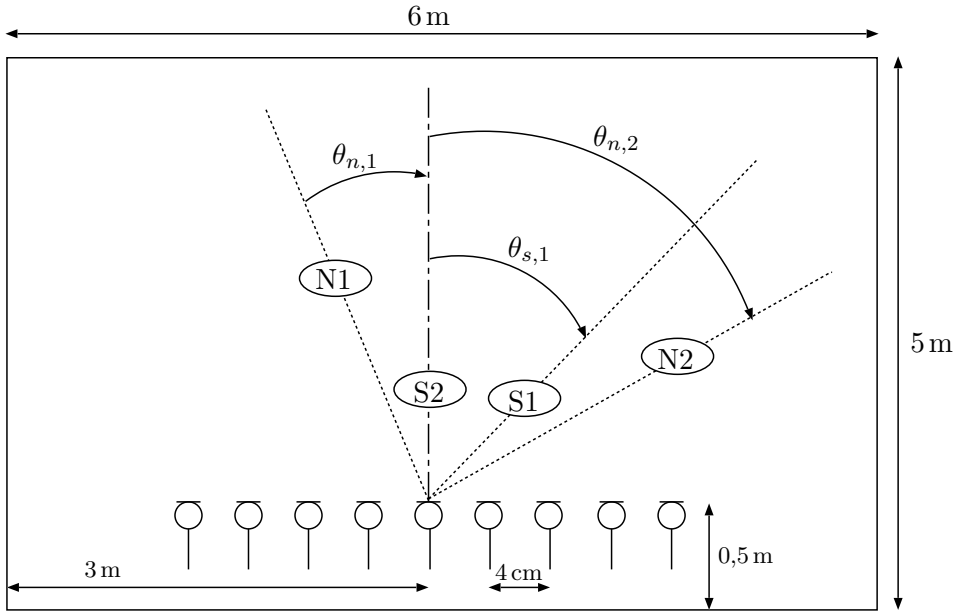

Bild C.1: Simulierte geometrische Anordnung für die Störgeräuschunterdrückung. Für die Nutzsignalquellen gilt ein radialer Abstand von 0,8 m und $\theta_{s,1} = 45°$, sowie $\theta_{s,2} = 0°$. Für die Störquellen gilt ein radialer Abstand von 1,6 m und $\theta_{n,1} = -20°$, sowie $\theta_{n,2} = 60°$.

Szenario-3 Sprachquelle S2 und Störquelle N2 sind aktiv

Szenario-4 Sprachquelle S2 und beide Störquellen N1 und N2 sind aktiv

Bei der Erzeugung der Mikrophonsignale mittels der Spiegelquellenmethode variiert die Nachhallzeit T_{60}, das SNR und die Anzahl der verwendeten Mikrophone. Diese Angaben sind jeweils an der Stelle in dieser Arbeit zu finden, an denen die Signale verwendet wurden.

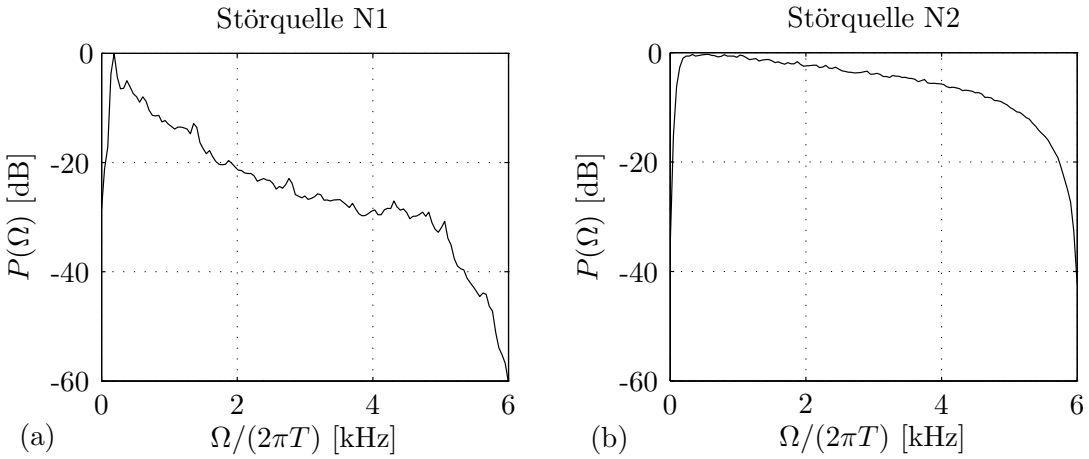

Bild C.2: Leistungsdichtespektrum $P(\Omega)$ in (a) für die Störquelle N1 und in (b) für für die Störquelle N2.

C.2 Spiegelquellenmethode für blinde Quellentrennung

Zur Untersuchung der Separationsleistung bei der blinden Quellentrennung mittels PCA *Beamforming* wurden in einem simulierten Raum mit einer Länge von 6 m, einer Breite von 5 m

und einer Höhe von 3 m zwei simultan aktive Sprachsignalquellen S1 und S2 platziert. Der Abstand der Quellen zum Mittelpunkt der linearen Mikrophongruppe beträgt jeweils 2 m und die Ausrichtungen betragen $\theta_{s,1} = -30\,^\circ$, sowie $\theta_{s,2} = 45\,^\circ$. Es wurden wiederum 10 Sprachbeispiele von 5 männlichen und 5 weiblichen Sprechern verwendet, wodurch sich insgesamt 45 Kombination ergeben. Dabei sind die beiden verhallten Signale mit gleicher Leistung an den Mikrophonen aufaddiert und zusätzlich unkorreliertes weißes Rauschen mit einem SNR von 25 dB hinzugefügt worden. Die Anordnung in dem simulierten Raum kann dem Bild C.3 entnommen werden. Im dieser Arbeit ist die Anordnung aus Bild C.3 mit *Szenario-5* bezeich-

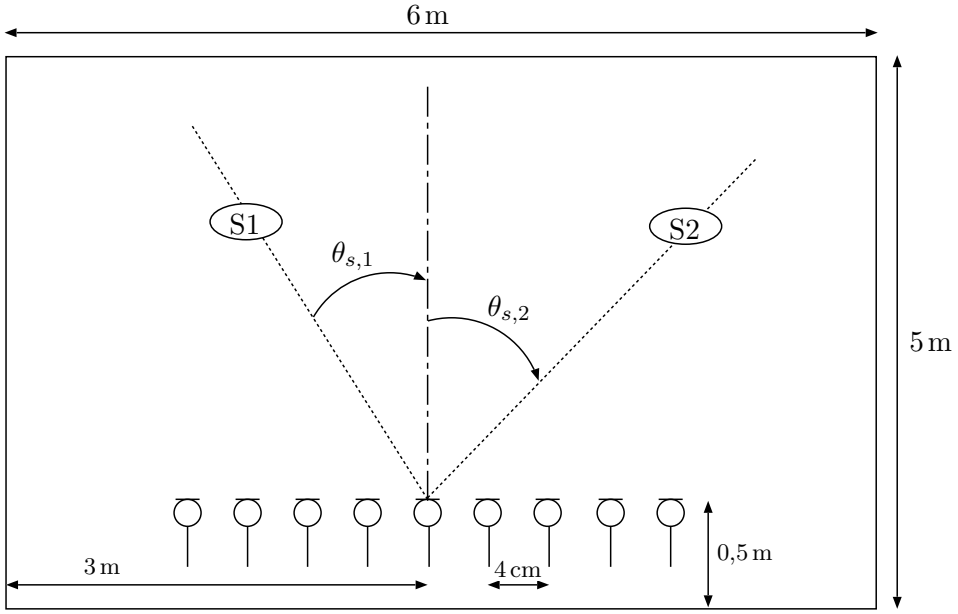

Bild C.3: Simulierte geometrische Anordnung für die blinde Quellentrennung. Die Sprachsignalquellen haben einen radialen Abstand von 2 m und die Einfallsrichtungen sind $\theta_{s,1} = -30\,^\circ$, sowie $\theta_{s,2} = 45\,^\circ$.

net. Die Verhallung wurde wieder mit der Spiegelquellenmethode durchgeführt, wobei die Nachhallzeit T_{60} und die Anzahl der verwendeten Mikrophone variiert wurden.

Anhang D

Robuste Sprache/Pause-Detektion

In Sprachsignalverarbeitungssystemen zur Telekommunikation oder zur akustischen Szenenanalyse ist die Detektion von Sprachaktivität eine sehr wichtige, fundamentale Komponente [SHU07]. Abhängig von der konkreten Anwendung sind unterschiedliche Strategien zur Sprache/Pause-Detektion (engl. *Voice Activity Detection*, VAD) notwendig. Bei z. B. der automatischen Spracherkennung müssen alle Segmente, welche Sprachanteile beinhalten vertrauenswürdig identifiziert werden und es sollte kein Sprachsegment ausgelassen werden [ETS02]. Beim Einsatz zur Schätzung von spektralen Leistungsdichten – wie hier in dieser Arbeit – ist es jedoch akzeptabel, nicht jedes Segment, sei es Sprache oder Pause, als solches zu identifizieren. Vielmehr sollte beim Entwurf darauf geachtet werden, dass wenn eine Klassifizierung als Sprache oder Pause erfolgt, diese auch sehr vertrauenswürdig ist. Daher soll eine VAD mit drei möglichen Klassen bzw. Zuständen eingesetzt werden: Zu den sonst üblichen Sprache und Pause Zuständen wird noch ein weiterer unentschiedener Zustand (engl. *don't know*) hinzugefügt.

Typischerweise kann das Klassifikationsproblem in zwei Teilen betrachtet werden: der Generierung von Entscheidungsmerkmalen und der Anwendung einer Entscheidungsregel. Als Entscheidungsmerkmal kann z. B. die Signalenergie dienen [SKS99, MK02, ETS02, WHUS07] oder die inhärente Charakteristik von Sprachsignalen [KDO05, Tuc92, IN06]. Basierend auf den generierten Merkmalen erfolgt dann die eigentliche Klassifikation z. B. mittels einer einfachen Schwellwertentscheidung oder statistisch motiviert über das Verhältnis von Wahscheinlichkeitsdichtefunktionen (engl. *Likelihood Ratio Test*, LRT). Im Folgenden soll die VAD nach [SKS99] analysiert und modifiziert werden. Hierbei dient die Signalenergie, oder genauer gesagt das SNR als Enscheidungsmerkmal und die Entscheidungsregel ist der *Likelihood Ratio Test*.

D.1 Likelihood-Ratio-Entscheidungsregel

Das einkanalige Mikrophonsignal $X(\Omega_k)$ soll im Frequenzbereich für jede diskrete Spektralkomponente Ω_k aus der Komponente des Sprachanteils $S(\Omega_k)$ und einem unkorrelierten additiven Rauschterm $N(\Omega_k)$ bestehen, wobei an dieser Stelle auf den Blockindex verzichtet werden soll. Weiterhin wird angenommen, dass der Sprach- und Rauschanteil jeweils komplexe Gaußverteilungen besitzt. Dann können die bedingten Wahrscheinlichkeitsdichtefunktionen $p(X(\Omega_k)|H_0(\Omega_k))$ bezüglich der Beobachtung einer spektralen Rauschkomponente gegeben

die Hypothese $H_0(\Omega_k)$ einer Sprachpause und entsprechend $p(X(\Omega_k)|H_1(\Omega_k))$ für die Beobachtung von Sprache und Rauschen gegeben die Hypothese $H_1(\Omega_k)$ für Sprachaktivität geschrieben werden als

$$p(X(\Omega_k)|H_0(\Omega_k)) = \frac{1}{\pi\sigma_N^2(\Omega_k)} \exp\left\{-\frac{|X(\Omega_k)|^2}{\sigma_N^2(\Omega_k)}\right\} \tag{D.1}$$

$$p(X(\Omega_k)|H_1(\Omega_k)) = \frac{1}{\pi(\sigma_N^2(\Omega_k)+\sigma_S^2(\Omega_k))} \exp\left\{-\frac{|X(\Omega_k)|^2}{\sigma_N^2(\Omega_k)+\sigma_S^2(\Omega_k)}\right\}, \tag{D.2}$$

wobei $\sigma_N^2(\Omega_k)$ und $\sigma_S^2(\Omega_k)$ die Varianzen von $N(\Omega_k)$ und $S(\Omega_k)$ bezeichnen. Das frequenzabhängige *Likelihood Ratio* ist definiert als

$$\Lambda(\Omega_k) = \frac{p(X(\Omega_k)|H_1(\Omega_k))}{p(X(\Omega_k)|H_0(\Omega_k))} = \frac{1}{1+\xi(\Omega_k)} \exp\left\{\frac{\gamma(\Omega_k)\xi(\Omega_k)}{1+\xi(\Omega_k)}\right\}, \tag{D.3}$$

mit dem so genannten *a posteriori* SNR

$$\gamma(\Omega_k) = \frac{|X(\Omega_k)|^2}{\sigma_N^2(\Omega_k)} \tag{D.4}$$

und dem *a priori* SNR

$$\xi(\Omega_k) = \frac{\sigma_S^2(\Omega_k)}{\sigma_N^2(\Omega_k)}. \tag{D.5}$$

Die Frequenzkomponenten sind als unabhängig untereinander anzusehen. Unter Berücksichtigung aller Frequenzkomponenten kann das *Likelihood Ratio* als Produkt über alle Frequenzen (D.3) und nach Logarithmieren als Summe über alle frequenzabhängigen *Likelihood Ratios* angegeben werden. Daraus folgt dann die gemittelte Entscheidungsregel

$$\log(\Lambda) = \frac{1}{L}\sum_{k=0}^{L-1} \log(\Lambda(\Omega_k)) \underset{H_0(\Omega_k)}{\overset{H_1(\Omega_k)}{\gtrless}} \eta, \tag{D.6}$$

mit der Länge L für die diskrete Fourier-Transformation und der Entscheidungsschwelle η.

Robustheitssteigerung der Entscheidungsregel

Da gerade am Ende einer Sprachsequenz sehr wenig Energie in dem Signal vorhanden ist, führt die direkte Anwendung von Gl. (D.6) häufig zu verfrühten Pause-Entscheidungen. Daher kann eine Verzögerung (engl. *Hang-Over*) abfallender Werte von Λ vorgenommen werden. In [SKS99] wird hierfür ein Verfahren basierend auf einem *Hidden Markov Modell* (HMM) und in [CK01] eine empirisch motivierte Glättung der *Likelihood Ratio* vorgeschlagen. Als Erweiterung der Verarbeitung von Einzelbeobachtungen und einer Nachverarbeitung mittels HMM oder Glättung ist in [RSB+05] alternativ die Ausnutzung von Mehrfachbeobachtungen in die *Likelihood*-Entscheidungsregel integriert. In zahlreichen Tests, welche im Rahmen dieser Arbeit durchgeführt wurden, hat sich die Glättung nach [CK01] als sehr effektive Variante herausgestellt:

$$\Psi_m(\Omega_k) = \exp\{\beta\log(\Psi_{m-1}(\Omega_k)) + (1-\beta)\log(\Lambda_m(\Omega_k))\}, \tag{D.7}$$

wobei nun der Blockindex m in der Rekursion Gl. (D.7) aufgeführt ist. Mit β ist die Glättungskonstante bezeichnet, die z. B. zu $\beta = 0.85$ gesetzt werden kann. Äquivalent zu Gl. (D.6) ergibt sich dann folgende Entscheidungsregel:

$$\log(\Psi_m) = \frac{1}{L}\sum_{k=0}^{L-1}\log(\Psi_m(\Omega_k)) \underset{H_0(\Omega_k)}{\overset{H_1(\Omega_k)}{\gtrless}} \eta. \tag{D.8}$$

D.2 Schätzung des a priori SNR

Um nun die Regel Gl. (D.8) auswerten zu können ist es notwendig das *a priori* SNR Gl. (D.5) für jeden Block m zu schätzen, z. B. mit Hilfe der so genannten *Decision-Directed* (DD) Methode nach [EM84]:

$$\hat{\xi}_m(\Omega_k) = \alpha\frac{\hat{S}^2_{m-1}(\Omega_k)}{\hat{\sigma}^2_{N,m-1}(\Omega_k)} + (1-\alpha)\mathrm{MAX}\{\gamma_m(\Omega_k), 1\}, \tag{D.9}$$

wobei $\hat{S}^2_m(\Omega_k)$ die geschätzte Amplitude der Sprache, α eine Glättungskonstante (z. B. $\alpha = 0.96$) und $\mathrm{MAX}\{\cdot\}$ der Maximum-Operator ist, mit $\mathrm{MAX}\{\psi, \vartheta\} = \psi$ für $\psi > \vartheta$, und sonst $\mathrm{MAX}\{\psi, \vartheta\} = \vartheta$. Der Amplitudenschätzer ergibt sich nach [EM84] zu

$$\hat{S}_m(\Omega_k) = \sqrt{\frac{\pi}{2}}\frac{\sqrt{\upsilon_m(\Omega_k)}}{\hat{\gamma}_m(\Omega_k)} M\left\{-0,5; 1; -\upsilon_m(\Omega_k)\right\} |X_m(\Omega_k)| \tag{D.10}$$

mit der konfluent hypergeometrischen Funktion

$$\begin{aligned} M\left\{-0,5; 1; -\upsilon_m(\Omega_k)\right\} = \exp\left\{-\frac{\upsilon_m(\Omega_k)}{2}\right\} \cdot \Bigg[&(1+\upsilon_m(\Omega_k))I_0\left\{\frac{\upsilon_m(\Omega_k)}{2}\right\} \\ &+\upsilon_m(\Omega_k)I_1\left\{\frac{\upsilon_m(\Omega_k)}{2}\right\}\Bigg], \end{aligned} \tag{D.11}$$

wobei

$$\upsilon_m(\Omega_k) = \frac{\hat{\xi}_m(\Omega_k)}{1+\hat{\xi}_m(\Omega_k)}\hat{\gamma}_m(\Omega_k). \tag{D.12}$$

Mit $I_0\{\cdot\}$ in Gl. (D.11) ist die modifizierte Besselfunktion nullter Ordnung und mit $I_1\{\cdot\}$ der ersten Ordnung bezeichnet. Da die Auswertung der Besselfunktionen sehr rechenintensiv ist wurde für die Implementierung der VAD folgende Approximation von Gl. (D.11) eingesetzt:

$$M\left\{-0,5; 1; -\upsilon\right\} \approx \widehat{M}(\upsilon) = 1,163\sqrt{\upsilon + 1,1} - 0,0015\upsilon - 0,22 \tag{D.13}$$

wobei in Gl. (D.13) auf den Frequenz- und Blockindex verzichtet wurde. In Bild D.1 (a) ist der Verlauf der hypergeometrischen Funktion für einen relevanten Wertebereich von υ dargestellt und in D.1 (b) das Quadrat des relativen Fehlers

$$e_r(\upsilon) = \frac{M\left\{-0,5; 1; -\upsilon\right\} - \widehat{M}(\upsilon)}{M\left\{-0,5; 1; -\upsilon\right\}}. \tag{D.14}$$

An Bild D.1 ist deutlich zu erkennen, dass Gl. (D.13) eine sehr gute Näherung darstellt.

Zur Berechnung des *a priori* SNRs Gl. (D.5) ist nun noch die Varianz des Rauschens zu schätzen. Dieses kann z. B. in den Sprachpausen erfolgen welches hier als implizite Schätzung bezeichnet werden soll oder es wird extern z. B. mit Hilfe der Minimum Statistik (MS) Methode nach [Mar01] berechnet, welches als explizite Schätzung bezeichnet werden soll.

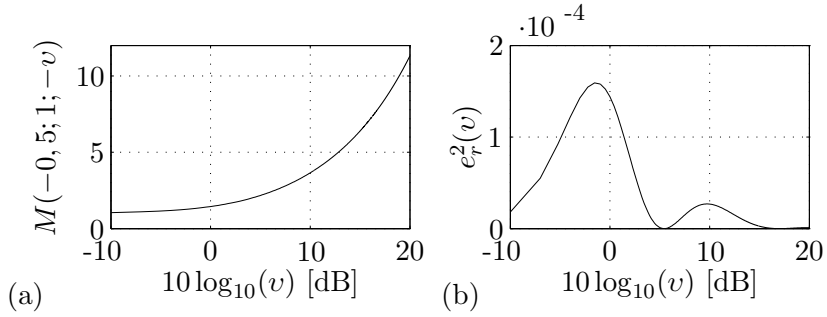

Bild D.1: Verlauf der konfluent hypergeometrischen Funktion nach Gl. (D.11) in (a) und in (b) der quadratische Fehler der Approximation nach Gl. (D.13).

Implizite Schätzung der Rauschvarianz

Das *Likelihood Ratio* soll als Informationsquelle zur Schätzung einer Sprachpause verwendet werden. Da dies für jede Frequenz erfolgt, wird im Folgenden auf den Frequenzindex erzichtet. Mit Hilfe der Bayes'schen Regel für bedingte Verteilungsdichtefunktionen $p(H_{0,m}|X_m)p(X_m) = p(X_m|H_{0,m})p(H_{0,m})$ und $p(X_m) = p(X_m|H_{0,m})p(H_{0,m}) + p(X_m|H_{1,m})p(H_{1,m})$ kann die Wahrscheinlichkeit für eine Sprachpause gegeben die Beobachtung X_m geschrieben werden als

$$p(H_{0,m}|X_m) = \frac{1}{1+\Upsilon_m} \tag{D.15}$$

mit

$$\Upsilon_m = \frac{p(H_{1,m})p(X_m|H_{1,m})}{p(H_{0,m})p(X_m|H_{0,m})} = \frac{p(H_{1,m})}{p(H_{0,m})}\Psi_m. \tag{D.16}$$

In Anlehnung an [SKS99] soll Υ_m rekursiv basierend auf einem *Hidden Markov Modell* berechnet werden. In dem benutzen zeitinvarianten Markov Prozess bezeichnet a_{ij} den Zustandsübergang von der Hypothese H_i nach H_j, mit $i,j \in \{1,2\}$. Die Werte sind empirisch gesetzt auf: $a_{00} = 0,8;\ a_{01} = 0,2;\ a_{10} = 0,1;\ a_{11} = 0,9$. Die Rekursionsgleichung für Gl. (D.16) ergibt sich dann zu:

$$\Upsilon_m = \frac{p(H_{0,m-1}, X_{m-1})a_{01} + p(H_{1,m-1}, X_{m-1})a_{11}}{p(H_{0,m-1}, X_{m-1})a_{00} + p(H_{1,m-1}, X_{m-1})a_{10}}\Psi_m \tag{D.17}$$

$$= \frac{a_{01} + \Upsilon_{m-1}a_{11}}{a_{00} + \Upsilon_{m-1}a_{10}}\Psi_m. \tag{D.18}$$

Die frequenzabhängige Rauschvarianz kann somit rekursiv berechnet werden zu

$$\hat{\sigma}^2_{N,m}(\Omega_k) = \alpha\hat{\sigma}^2_{N,m-1}(\Omega_k) + (1-\alpha)E\{|N_m(\Omega_k)|^2|X_m(\Omega_k)\} \tag{D.19}$$

mit

$$\begin{aligned} E\{|N_m(\Omega_k)|^2|X_m(\Omega_k)\} \approx\ & p(H_{0,m}(\Omega_k)|X_m(\Omega_k))|X_m(\Omega_k)|^2 \\ & + (1 - p(H_{0,m}(\Omega_k)|X_m(\Omega_k)))\hat{\sigma}^2_{N,m-1}(\Omega_k), \end{aligned} \tag{D.20}$$

wobei $p(H_{0,m}(\Omega_k)|X_m(\Omega_k))$ in Gl. (D.20) aus Gl. (D.15) durch Einsetzen von Gl. (D.18) hervorgeht.

Explizite Schätzung der Rauschvarianz

Die Grundidee der Minimum Statistik nach [Mar94] besteht darin, dass das Minimum der spektralen Leistungsdichte auf das zu schätzende Rauschen zurückzuführen ist. Dieses kann folglich durch eine Minima-Suche in einer gewissen Anzahl von vergangenen Verarbeitungsblöcken pro Spektralkomponente auch während Sprachaktivität ermittelt werden. Offensichtlich besteht jedoch zwischen der zu schätzenden Rauschvarianz und den so bestimmten Minima eine systematische Fehlschätzung. Daher wurde in [Mar01] ein Verzerrungsfaktor als Korrekturterm eingeführt. Aufgrund der Komplexität des Verfahrens sei auf [Mar01] für weitere Details verwiesen. An dieser Stelle soll lediglich die Fähigkeit des implementierten Algorithmus, eine kontinuierliche Schätzung der Störgeräuschleistung auch während Sprachsequenzen durchzuführen, anhand des Bildes D.2 exemplarisch verdeutlicht werden. Auf der gesamten Länge des ausgewählten Zeitintervalls liegt Sprachaktivität vor und dem Sprachsignal wurde weißes Rauschen mit zeitvarianter Leistungsdichte in Form zweier Sägezähne überlagert. Das SNR variert in dem Bereich zwischen 0 dB und 15 dB. In Bild D.2 ist zum einen das geglättete Periodogramm der resultierenden Spektralkomponente bei ca. 1 kHz und zum anderen das geschätzte Störspektrum $B_c \cdot P_{\min}$ über der Zeit aufgetragen. Hierbei bezeichnet $P_{\min}$ das ermittelte Minimum und B_c den Korrekturterm. Ohne quantitative Aussagen zu treffen ist in Bild D.2 rein qualitativ zu erkennen, dass die Schätzung der Störung dem Sägezahnverlauf folgt.

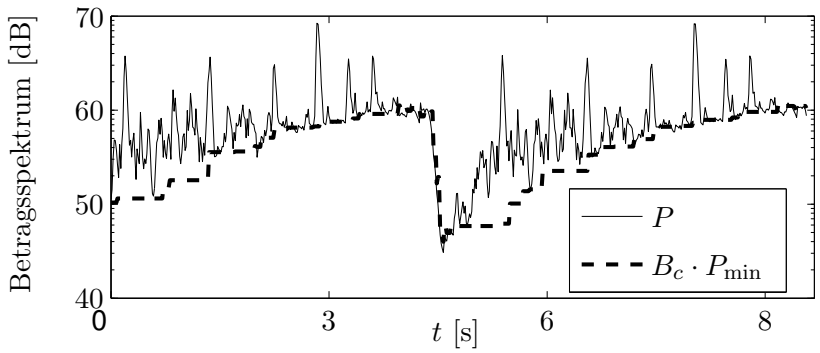

Bild D.2: Exemplarische Darstellung der Schätzung der Rauschvarianz nach [Mar01] für eine Spektralkomponente des Sprachsignals bei 1 kHz, welches mit einem sägezahnförmigen Rauschen in dem Bereich zwischen 0 dB und 15 dB überlagert wurde.

D.3 Analyse von Fehlschätzungen der Rauschvarianz

Die relative Abweichung des *Likelihood Ratios* soll in Abhängigkeit von einer Fehlschätzung der Rauschvarianz untersucht werden, zuerst für eine Überschätzung der Varianz, einerseits verursacht durch Einbeziehung von Sprachanteilen in die Schätzung aber andererseits auch durch zeitliche Änderungen der Rauschstatistik. Danach erfolgt eine Analyse für eine Unterschätzung der Rauschvarianz. Da das prinzipielle Verhalten für alle Frequenzen gleich ist wird wieder auf den Frequenzindex verzichtet. Die Abweichung wird nun zuerst definiert zu

$$\Delta\sigma_N^2 = K_S \sigma_S^2, \tag{D.21}$$

wobei der Koeffizient $K_S \in [0, .., 1]$ die Größe der Abweichung relativ zur Varianz der Sprache angibt. Dann kann das *a priori* SNR angegeben werden als

$$\tilde{\xi} = \frac{\sigma_S^2}{\sigma_N^2 + \Delta\sigma_N^2} = \frac{1}{\xi^{-1} + K_S}, \tag{D.22}$$

mit dem wahren *a priori* SNR $\xi = \sigma_S^2/\sigma_N^2$. Es soll angenommen werden, dass das *a posteriori* SNR gegeben ist durch $\gamma = \xi + 1$, wodurch sich die *Likelihood-Ratio*-Abweichung angeben läßt zu

$$\Delta \log(\Lambda) = \left(\frac{\gamma\xi}{1+\xi} - \log(1+\xi) \right) - \left(\frac{\gamma\tilde{\xi}}{1+\tilde{\xi}} - \log(1+\tilde{\xi}) \right). \tag{D.23}$$

Nimmt man nun ein bestimmtes *a priori* SNR an, so kann die Erhöhung des *Likelihood Ratios* $\Delta \log(\Lambda)$ für unterschiedliche Abweichungen $\Delta\sigma_N^2$ berechnet werden.

Einen etwas anderen Ausdruck für Gl. (D.22) erhält man, wenn die Abweichung der geschätzten Rauschvarianz angenommen wird zu

$$\Delta\sigma_N^2 = K_N \sigma_N^2, \tag{D.24}$$

wobei der Koeffizient $K_N \in]-1, .., 0]$ nun die Größe der Abweichung relativ zur Rauschvarianz festlegt. Mit dieser Differenz ergibt sich dann das *a priori* SNR

$$\tilde{\xi} = \frac{\xi}{1 + K_N}, \tag{D.25}$$

welches wiederum in Gl. (D.23) eingesetzt werden kann.

In Bild D.3 ist Gl. (D.23) exemplarisch ausgewertet für die fehlerhaft geschätzten *a priori* SNR nach Gl. (D.22) und Gl. (D.25). Bild D.3 zeigt offensichtlich ein sehr sensibles Verhalten

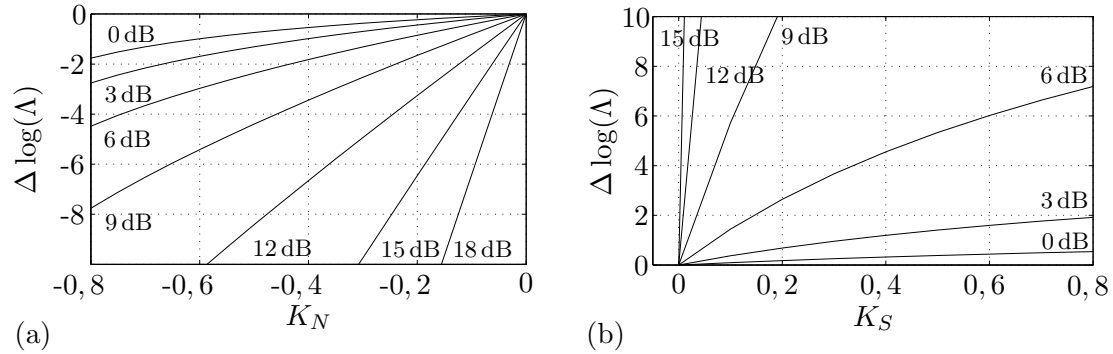

Bild D.3: Abweichung des *Likelihood Ratios* nach Gl. (D.23) für unterschiedliche *a priori* SNR: In (a) relativ zur Varianz des Rauschens ($\Delta\sigma_N^2 = K_N\sigma_N^2$) und in (b) relativ zur Varianz der Sprache ($\Delta\sigma_N^2 = K_S\sigma_S^2$).

der Entscheidungsregel bezüglich der Schätzung der Rauschvarianz. Daher ist es zwingend notwendig, einerseits ein möglichst schnelles Nachführen von $\hat{\sigma}_{N,m}^2(\Omega_k)$ zu ermöglichen, aber andererseits sicherzustellen, dass keine Sprachanteile in die Schätzung einfließen. Das Verhalten der Entscheidungsregel bezüglich der Abbildung D.3 (b) kann weitergehend dahin interpretiert werden, dass falls Energie der Sprache in die Schätzung der Varianz des Rauschens einfließt, der Wert $\log(\Lambda(m))$ sprunghaft ansteigt und somit noch sicherer Sprachpausen detektiert werden. Somit erfolgt dann wieder eine zuverlässige Rückführung von $\hat{\sigma}_{N,m}^2(\Omega_k)$ auf den wahren Wert.

Robustheitssteigerung der Rauschvarianzschätzung

Insbesondere beim Einsetzen von Sprache bzw. beim Ausklingen ist $p(H_{0,m}(\Omega_k)|X_m(\Omega_k))$ in Gl. (D.20) eventuell nicht schnell genug nachgeführt. Um nun ein Lecken von Sprachanteilen in die Schätzung von $\hat{\sigma}^2_{N,m}(\Omega_k)$ zu verhindern wird eine Hintergrundschätzung von $E\{|N_m(\Omega_k)|^2|X_m(\Omega_k)\}$ in Gl. (D.20) in einem Schieberegister vorgenommen und die Werte werden erst in Gl. (D.19) verwendet, wenn z. B. in 10 aufeinanderfolgenden Blöcken $p(H_{0,m}(\Omega_k)|X_m(\Omega_k)) > 0,2$ gilt. Allerdings beginnt das Füllen des Registers erst nach einem gewissen Offset von z. B. 20 aufeinanderfolgenden Blöcken mit $p(H_{0,m}(\Omega_k)|X_m(\Omega_k)) > 0,2$.

D.4 Simulationen

Es sollen nun experimentelle Ergebnisse für die Detektionsgenauigkeit der VAD folgen. Dafür wurden 20 Äußerungen von verschiedenen Sprechern (10 männlich und 10 weiblich, abgetastet mit 12 kHz) zu einem Audiosignal der Länge 120 Sekunden mit einem Sprachanteil von ungefähr 50 % zusammengefaßt. Eine manuelle Markierung des reinen Sprachsignals auf Verarbeitungsblöcken der Länge 128 diente als Referenz für die Auswertungen. Die DFT-Länge der VAD wurde auf $L = 256$ gesetzt, wobei jeweils sich halb überlappende Blöcke nach einer Hamming-Fensterung transformiert wurden.

Stationäres Rauschen

Dem reinen Signal wurde nun stationäres weißes Rauschen mit unterschiedlichem SNR im Bereich von 0 dB bis 25 dB überlagert. In der Signalentdeckungstheorie stellt die *Receiver Operating Characteristic* (ROC) Kurve eine Methode zur Darstellung von Fehlern binärer Entscheidungen dar und dient der Grenzwertoptimierung. Man ermittelt für jeden möglichen Grenzwert – hier die Entscheidungsvariable η – die resultierenden relativen Häufigkeitsverteilungen und errechnet die jeweils zugehörige Sensitivität und Spezifität. Im Diagramm gibt die Ordinate die Sensitivität (= relative Häufigkeit aller richtig-positiven Testergebnisse) und die Abszisse die Spezifität (= relative Häufigkeit aller falsch-positiven Testergebnisse) an. Im Falle der VAD bezeichnet die Sensitivität die Fälle $p(\log(\Psi_m) > \eta|H_{m,1})$ und die Spezifität die Fälle $p(\log(\Psi_m) > \eta|H_{m,0})$. Die resultierenden ROC Kurven sind in Bild D.4 dargestellt. Es ist sehr deutlich die hohe Detektionsgenauigkeit insbesondere für mittlere SNR-Werte zu erkennen.

Robustheitssteigerung der Detektionsgenauigkeit

Da jedoch die Werte $\log(\Psi_m)$ für Sprache und Pause bei niedrigen SNR deutlich enger beieinander liegen als für hohe SNR, ist eine gute Wahl für den Arbeitspunkt der Entscheidungsvariablen η nicht für einen großen Dynamikbereich der erwarteten SNR möglich. Daher ist es sinvoll zwei Schwellwerte η_0 und η_1, mit $\eta_0 < \eta_1$, einzuführen und eine Pause anzuzeigen, wenn gilt $\log(\Psi_m) < \eta_0$ bzw. Sprache anzuzeigen für $\log(\Psi_m) > \eta_1$. Daraus folgt, dass für $\eta_0 \leq \log(\Psi_m) \leq \eta_1$ der unentschiedene Zustand eintritt. Da für die Anwendung der VAD in dieser Arbeit zwar eine sichere Detektion von Sprachsegmenten erforderlich und aber gleichzeitig ein schnelles Nachführen der entsprechenden Algorithmen bei Sprachaktivität wünschenswert ist, wurde $\eta_1 = 0,8$ aus den Auswertungen der Simulationen gewählt. Unter der Annahme einer stationären Störung, bzw. einer sich nur sehr langsam ändernden

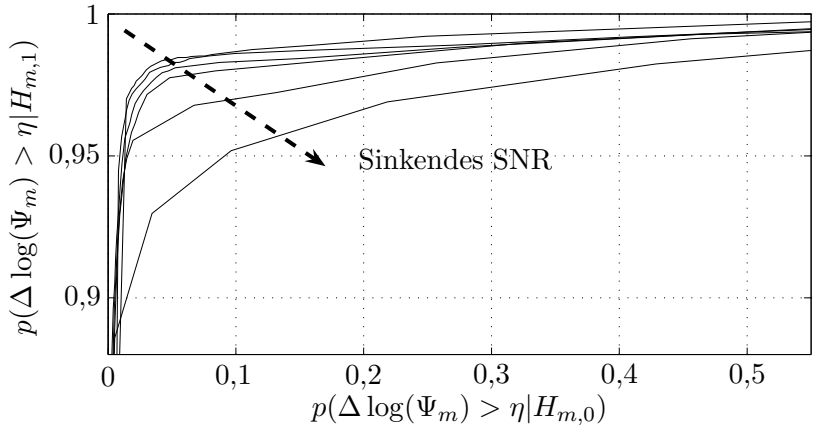

Bild D.4: ROC-Kurven für stationäres weißes Rauschen mit unterschiedlichem SNR: 25 dB, 20 dB, 15 dB, 10 dB, 5 dB und 0 dB. Schätzung der Rauschvarianz mit impliziter Methode nach Gl. (D.19) und Gl. (D.20) unter Beachtung der aufgeführten Robustheitsaspekte.

Rauschstatistik, kann ein Verpassen von Pausesegmenten sehr wohl geduldet werden, wodurch umgekehrt bei der Detektion von Pausen diese auch mit einer höheren Wahrscheinlichkeit korrekt sind. Daher wurde $\eta_0 = 0,2$ gewählt. Die sich ergebenden Detektionsgenauigkeiten sind in der Tabelle D.1 zusammengefaßt.

	Sprache		Pause	
	falsch	korrekt	falsch	korrekt
SNR	$p(\log(\Psi_m) > \eta_1 \mid H_{m,0})$	$p(\log(\Psi_m) > \eta_1 \mid H_{m,1})$	$p(\log(\Psi_m) < \eta_0 \mid H_{m,1})$	$p(\log(\Psi_m) < \eta_0 \mid H_{m,0})$
0 dB	0,02 %	69,00 %	1,76 %	53,44 %
5 dB	0,35 %	87,41 %	1,10 %	54,41 %
10 dB	1,14 %	93,88 %	1,03 %	57,12 %
15 dB	2,45 %	97,10 %	0,87 %	59,75 %
20 dB	3,65 %	98,04 %	0,71 %	67,56 %
25 dB	4,97 %	98,40 %	0,52 %	73,05 %

Tabelle D.1: Detektionsergbnisse für falsch bzw. korrekt detektierte Sprache- und Pause-Segmente unter Verwendung der VAD mit drei Zuständen für variierendes SNR.

Instationäres Rauschen

Als nächstes sollen noch ROC-Kurven präsentiert werden für einen Vergleich der impliziten Schätzung der Rauschvarianz nach Gl. (D.20) und der expliziten kontinuierlichen Schätzung mit Hilfe des Minimum-Statistik-Verfahrens, jeweils eingesetzt in Gl. (D.19). Dieser Test wurde für drei Arten von Rauschszenarien durchgeführt: Stationäres weißes Rauschen mit einem SNR von 10 dB, für sich sprunghaft änderndes weißes Rauschen zwischen einem SNR von 10 dB und 20 dB (siehe Bild D.5 (a)) und für sich pulsierend änderndes weißes Rauschen im Bereich zwischen einem SNR von 6 dB und 14 dB (siehe Bild D.5 (b)). Die Ergebnisse für die drei Rauschszenarien sind in Bild D.6 dargestellt. Zum einen ist in der Abbildung zu sehen, dass bei stationärem Rauschen die Ergebnisse mit der expliziten Schätzung minimal schlechter sind als mit der impliziten Methode. Dies ist durch die kontinuierliche Schätzung der Minimum-Statistik-Methode zu erklären, da so stets kleine Änderungen der Rauschvarianz

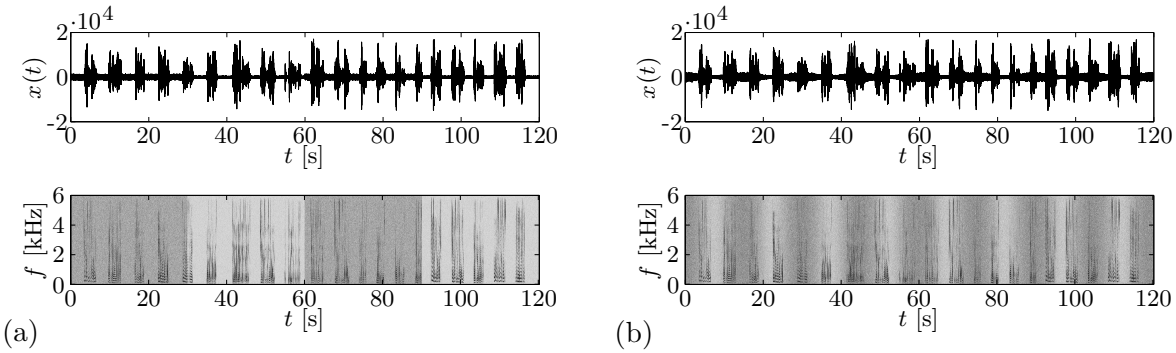

Bild D.5: Zeitverläufe und Spektrogramme der beiden verwendeten nichtstationären Rauscharten: In (a) sprunghafte Änderng des Rauschens zwischen einem SNR von 10 dB und 20 dB und in (b) pulsierendes Rauschen im Bereich zwischen einem SNR von 6 dB und 14 dB.

über der Zeit auftreten, die sich aber negativ auf die Entscheidungsregel auswirken. Zum anderen wird deutlich, dass mit der impliziten Schätzmethode bei instationärem Rauschen keine zuverlässigen Sprachaktivitätsentscheidungen mehr zu treffen sind. Hingegen liefert die VAD betrieben mit der expliziten Rauschschätzung weiterhin akzeptable Ergebnisse.

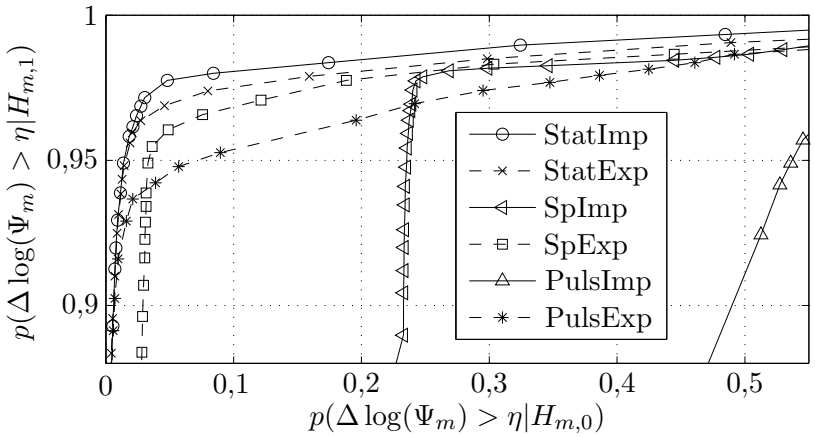

Bild D.6: ROC-Kurven für stationäres weißes Rauschen (bezeichnet mit "Stat"), für sich sprunghaft änderndes weißes Rauschen (bezeichnet mit "Sp") und für sich pulsierend änderndes weißes Rauschen (bezeichnet mit "Puls"); jeweils für die implizite Schätzung der Rauschvarianz (bezeichnet mit "Imp") und expliziter Schätzung (bezeichnet mit "Exp").

D.5 Zusammenfassung

Das hier beschriebene Verfahren zur Sprache/Pause-Detektion erlaubt eine robuste Steuerung der im Verlauf dieser Arbeit vorgestellten *Beamforming*-Algorithmen. Da die Problemstellung bei der mehrkanaligen Sprachsignalverbesserung in der Unterdrückung stationärer Störgeräuschquellen lag, wird die VAD mit der impliziten Rauschvarianzschätzung betrieben. Weil die Analyse von Fehlschätzungen der Rauschvarianz ergeben hat, dass die Entscheidungsregel ein sehr sensitives Verhalten bezüglich Abweichungen der Schätzung aufweist, wurde zur Steigerung der Zuverlässigkeit der Rauschvarianzschätzung die beschriebene Hintergrundschätzung angewendet. Bei der Implementierung wurde insbesondere auf Robustheitsaspekte geachtet, die ein sicheres Erkennen von Pause- und Sprache-Segmenten gewährleisten. Diese

kamen bei der Glättung der Entscheidungsregel und insbesondere durch die Nutzung von drei Zuständen für die Klassifikation zum Tragen.

Anhang E

Adaptive Eigenwertzerlegung

In diesem Abschnitt soll zuerst die Originalherleitung der Oja-Regel präsentiert werden. Dann folgen experimentelle Ergebnisse für die Schrittweite von Gradientenverfahren zur Lösung des speziellen und des allgemeinen Eigenwertproblems, welche essentiell für die Stabilität der Algorithmen ist.

E.1 Oja Lernregel

Die Originalherleitung der Oja-Regel nach [Oja82] basiert auf einer Normierung der Filterkoeffizienten und der anschließenden Taylorreihenentwicklung, also ohne den Ansatz mittels Lagrange-Multiplikator, wobei $C = 1$ gewählt ist. Die Maximierungsaufgabe ist nachwievor Gl. (5.26) und normiert wird nun die Hebbsche Lernregel Gl. (5.21)

$$\hat{\mathbf{v}}_{1,\kappa} = \frac{\hat{\mathbf{v}}_{1,\kappa-1} + \mu \mathbf{X}_\kappa Y_\kappa^*}{||\hat{\mathbf{v}}_{1,\kappa-1} + \mu \mathbf{X}_\kappa Y_\kappa^*||}. \tag{E.1}$$

Mit der Vektornotation $\hat{\mathbf{v}}_{1,\kappa} = [\hat{v}_{1,1,\kappa}, \ldots, \hat{v}_{1,M,\kappa}]^T$ für die M Komponenten ergibt sich für den Nenner von Gl. (E.1) eine Funktion $f(\mu)$ abhängig von der Schrittweite

$$f(\mu) = \left(\sum_{i=1}^{M} [\hat{v}_{1,i,\kappa-1} + \mu Y_\kappa^* X_{i,\kappa}][\hat{v}_{1,i,\kappa-1}^* + \mu Y_\kappa X_{i,\kappa}^*] \right)^{1/2}. \tag{E.2}$$

Die Funktion $f(\mu)$ wird mittels Taylor-Entwicklung in der Umgebung des Punktes $\mu = \mu_0 = 0$ durch eine Potenzreihe $P_f(\mu_0)$ dargestellt

$$P_f(\mu_0) = \left[\sum_{i=1}^{M} |\hat{v}_{1,i,\kappa-1}|^2 \right]^{1/2} \left[1 + \frac{\mu}{2} \sum_{i=1}^{M} \left(Y_\kappa^* X_{i,\kappa} \hat{v}_{1,i,\kappa-1}^* + Y_\kappa X_{i,\kappa}^* \hat{v}_{1,i,\kappa-1} \right) \right] + \mathcal{R}(\mu^2) \tag{E.3}$$

$$= 1 + \mu Y_\kappa Y_\kappa^* + \mathcal{R}(\mu^2), \tag{E.4}$$

wobei $\mathcal{R}(\mu^2)$ die Restglieder zweiter und höherer Ordnung beschreibt, $Y_\kappa = \sum_{i=1}^{M} X_{i,\kappa} \hat{v}_{1,i,\kappa-1}^*$ gilt und die Nebenbedingung eingehalten sein soll ($||\hat{\mathbf{v}}_{1,\kappa-1}|| = 1$). Mit der Näherung

$$\frac{1}{1+\varepsilon} \approx 1 - \varepsilon \tag{E.5}$$

für ε nahe Null folgt nach Einsetzen von Gl. (E.4) in Gl. (E.1) mit Gl. (E.5)

$$\hat{\mathbf{v}}_{1,\kappa} = \left(\hat{\mathbf{v}}_{1,\kappa-1} + \mu \mathbf{X}_\kappa Y_\kappa^*\right)\left(1 - \mu Y_\kappa Y_\kappa^* - \mathcal{R}(\mu^2)\right). \tag{E.6}$$

Nach der Ausmultiplikation von Gl. (E.6) und dem Weglassen aller Terme der Ordnung $\mathcal{O}(\mu^2)$ bzw. höherer Ordnung ergibt sich letztendlich das selbe Ergebnis wie in Gl. (5.32)

$$\hat{\mathbf{v}}_{1,\kappa} = \hat{\mathbf{v}}_{1,\kappa-1} + \mu Y_\kappa^* (\mathbf{X}_\kappa - Y_\kappa \hat{\mathbf{v}}_{1,\kappa-1}). \tag{E.7}$$

E.2 Schrittweite

Ein wesentliches Problem von Gradientenverfahren ist die Wahl einer geeigneten Schrittweite. Wird diese klein gewählt, so ist die Konvergenzgeschwindigkeit gering, dafür sind aber auch die Schwankungen um den stationären Punkt klein. Möchte man allerdings eine schnelle Adaption realisieren ist die Schrittweite zwangsläufig auf einen möglichst hohen Wert zu setzen. Hierbei ist dann insbesondere darauf zu achten, dass das Gradientenverfahren nicht divergiert. Es ist also eine Abschätzung für eine maximale Schrittweite notwendig. Dies soll anhand von Simulationen zuerst für das spezielle und danach für das allgemeine Eigenwertproblem erfolgen.

Spezielles Eigenwertproblem

Es soll nun anhand von Simulationen die Stabilität der Oja-Regel für unterschiedliche Werte der Schrittweite untersucht und mit dem neuen Verfahren verglichen werden. Daher sollen die deterministischen Verfahren Gl. (5.31) und Gl. (5.36) durch hochgestellte Bezeichnung "(Oja)" und "(Neu)" an den Schrittweiten gekennzeichnet sein

$$\hat{\mathbf{v}}_{1,\kappa} = \begin{cases} \hat{\mathbf{v}}_{1,\kappa-1} + \mu^{(\mathrm{Oja})} \left(\boldsymbol{\Phi}_{\mathbf{XX}} - \hat{\mathbf{v}}_{1,\kappa-1}^H \boldsymbol{\Phi}_{\mathbf{XX}} \hat{\mathbf{v}}_{1,\kappa-1}\right) \hat{\mathbf{v}}_{1,\kappa-1}, & \text{Ojas Regel} \\ \dfrac{1 + \hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1}}{2\hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1}} \hat{\mathbf{v}}_{1,\kappa-1} + \mu^{(\mathrm{Neu})} \left(\boldsymbol{\Phi}_{\mathbf{XX}} - \dfrac{\hat{\mathbf{v}}_{1,\kappa-1}^H \boldsymbol{\Phi}_{\mathbf{XX}} \hat{\mathbf{v}}_{1,\kappa-1}}{\hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{v}}_{1,\kappa-1}}\right) \hat{\mathbf{v}}_{1,\kappa-1}, & \text{Neue Regel.} \end{cases} \tag{E.8}$$

Jeder Koeffizientenvektor $\hat{\mathbf{v}}_{1,\kappa}$ soll nun für jeden Iterationsschritt durch die Linearkombination der Eigenvektoren ausgedrückt werden

$$\hat{\mathbf{v}}_{1,\kappa} = \sum_{i=1}^{M} c_{i,\kappa} \mathbf{v}_i, \tag{E.9}$$

wobei $c_{i,\kappa}$ das Gewicht für den Iterationsschritt κ bezeichnet. Mit Gl. (E.9) wird aus Gl. (E.8)

$$\mathbf{c}_\kappa = \begin{cases} \mathbf{c}_{\kappa-1} + \mu^{(\mathrm{Oja})} \left(\boldsymbol{\Lambda} - \mathrm{diag}\left\{\mathbf{c}_{\kappa-1}^{\mathrm{H}} \boldsymbol{\Lambda} \mathbf{c}_{\kappa-1}\right\}\right) \mathbf{c}_{\kappa-1}, & \text{Ojas Regel} \\ \mathbf{c}_{\kappa-1} \dfrac{1 + \mathbf{c}_{\kappa-1}^H \mathbf{c}_{\kappa-1}}{2\mathbf{c}_{\kappa-1}^H \mathbf{c}_{\kappa-1}} + \mu^{(\mathrm{Neu})} \left(\boldsymbol{\Lambda} - \mathrm{diag}\left\{\dfrac{\mathbf{c}_{\kappa-1}^{\mathrm{H}} \boldsymbol{\Lambda} \mathbf{c}_{\kappa-1}}{\mathbf{c}_{\kappa-1}^{\mathrm{H}} \mathbf{c}_{\kappa-1}}\right\}\right) \mathbf{c}_{\kappa-1}, & \text{Neue Regel.} \end{cases} \tag{E.10}$$

Mit der Vektornotation $\mathbf{c}_\kappa = (c_{1,\kappa}, \ldots, c_{M,\kappa})^T$ und der Diagonalmatrix der Eigenwerte $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_\mathrm{i}\}$, welche der Größe nach angeordnet sein sollen $\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_M > 0$. Bei Ausführung der Rekursionsvorschriften Gl. (E.10) verschwinden die $c_{i,\kappa}$ mit $i > 1$ für große κ.

Dieses Verhalten wird nun als Funktion des Quadrats der Norm $\mathbf{c}_0^H \mathbf{c}_0 = K$ bei der Initialisierung betrachtet. Ein weiterer betrachteter Parameter ist das Verhältnis zwischen größtem und kleinstem Eigenwert $\chi = \lambda_1/\lambda_M$.

Durch Simulationen hat sich folgende Schreibweise zur Formulierung einer oberen Grenze $\mu_{\max}$ für die Schrittweite als geeignet erwiesen

$$\mu_{\max} = \frac{2}{\xi_{\min} \cdot \lambda_1}, \tag{E.11}$$

wobei experimentell unterschiedliche Werte $\xi_{\min}$ für Ojas Regel ($\xi_{\min}^{(\mathrm{Oja})}$) und für die neue Regel ($\xi_{\min}^{(\mathrm{Neu})}$) ermittelt wurden:

$$\xi_{\min}^{(\mathrm{Oja})} < \xi^{(\mathrm{Oja})}(\chi, K) = 1 + \frac{K-1}{2}(1 + \frac{1}{\chi}) < K \tag{E.12}$$

$$\xi_{\min}^{(\mathrm{Neu})} < \xi^{(\mathrm{Neu})}(\chi) = 1 - \frac{1}{\chi} < 1. \tag{E.13}$$

Beispielhafte Simulationsergebnisse für $\xi_{\min}^{(\mathrm{Oja})}$ und $\xi_{\min}^{(\mathrm{Neu})}$ sind in Bild E.1 für $K = 50$ und $K = 100$ dargestellt, wobei die Dimension $M = 4$ gewählt wurde. Zu sehen sind die markierten Messwerte, die gerade noch zu einer Konvergenz von Gl. (E.10) führen: für $K = 50$ markiert durch "x" und für $K = 100$ markiert durch "□". Außerdem sind die kontinuierlichen Verläufe der Funktionen Gl. (E.12) und Gl. (E.13) aufgetragen.

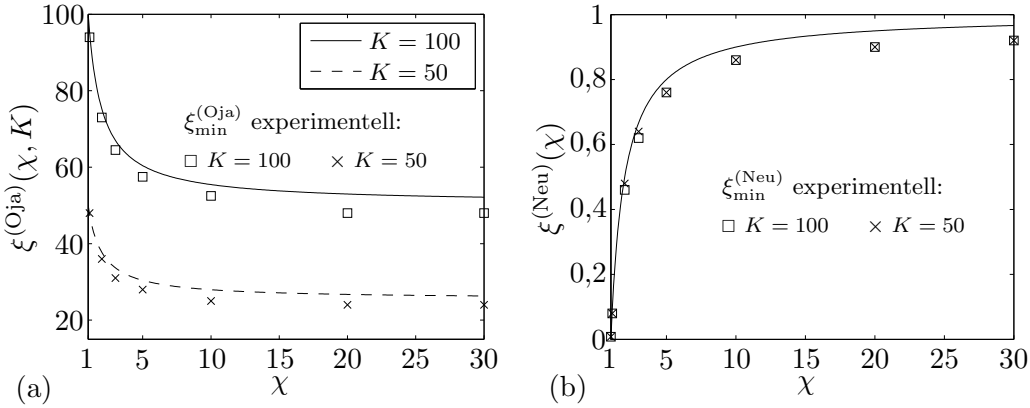

Bild E.1: Simulationsergebnisse der unteren Schranken $\xi_{\min}^{(\mathrm{Oja})}$ und $\xi_{\min}^{(\mathrm{Neu})}$ sowie der Verlauf der Abschätzungen $\xi^{(\mathrm{Oja})}(\chi, K)$ und $\xi^{(\mathrm{Neu})}(\chi)$ aus Gl. (E.12) und Gl. (E.13) für Ojas Regel in (a) und die neue Regel in (b).

Da eine temporäre, starke Abweichung der Norm des Vektors $\hat{\mathbf{v}}_{1,\kappa}$ von der Nebenbedingung unvorhersehbar ist und im fortlaufenden Betrieb durchaus vorkommen kann, ist die Unabhängigkeit der maximalen Schrittweite von der Norm $||\hat{\mathbf{v}}_{1,\kappa}||$ der neuen Regel sehr wünschenswert (vgl. Gl. (E.13) unabhängig von K). Andernfalls muss bei der direkten Verwendung der Oja-Regel die Schrittweite um eine Abschätzung für eine maximale Abweichung $K - 1$ reduziert werden.

Allgemeines Eigenwertproblem

Die beiden Varianten Algorithmus 9 (A-Grad-GG)/(A-RQgrad-GG) des Gradientenverfahrens sind nicht in eine Form äquivalent zu E.10 zu überführen. Daher wird die Schreibweise

$$\mu_\kappa = \frac{\rho}{r_\kappa} \tag{E.14}$$

für eine experimentelle Ermittlung der maximalen Schrittweite gewählt. Der Parameter r_κ stellt den Rayleigh Quotienten zum aktuellen Iterationsschritt dar. Der Faktor ρ wird nun auf stetig steigende Werte gesetzt bis schließlich die beiden Varianten des Gradientenverfahrens Gl. (E.16) nicht mehr konvergieren sondern divergieren. Mit diesem maximalen Wert $\rho_{\max}$ ergibt sich die maximale Schrittweite

$$\mu_{\max,\kappa} = \frac{\rho_{\max}}{r_\kappa}. \tag{E.15}$$

Die Experimente wurden mit akustischen Daten nach *Szenario-2* durchgeführt. Das gerichtete Tiefpassrauschen ist mit einem SNR von 5 dB dem 5-kanaligen Sprachsignal überlagert, und zusätzlich ist unkorreliertes Rauschen mit einem SNR von 25 dB hinzugefügt worden. Es werden die deterministischen Gradientenverfahren hergenommen mit perfekt bestimmten KLDS-Matrizen $\hat{\mathbf{\Phi}}_{\mathbf{XX}}$ und $\widetilde{\mathbf{\Phi}}_{\mathbf{NN}} = \hat{\mathbf{\Phi}}_{\mathbf{NN}}/\hat{\sigma}_N^2$, und $\hat{\sigma}_N^2 = \text{Spur}\{\hat{\mathbf{\Phi}}_{\mathbf{NN}}\}/\text{M}$:

$$\hat{\mathbf{v}}_{1,\kappa} = \frac{C^2 + \hat{\mathbf{v}}_{1,\kappa-1}^H \widetilde{\mathbf{\Phi}}_{\mathbf{NN}} \hat{\mathbf{v}}_{1,\kappa-1}}{2\hat{\mathbf{v}}_{1,\kappa-1}^H \widetilde{\mathbf{\Phi}}_{\mathbf{NN}} \hat{\mathbf{v}}_{1,\kappa-1}} \hat{\mathbf{v}}_{1,\kappa-1} + \begin{cases} \mu_\kappa^{(r)} \left(\hat{\mathbf{\Phi}}_{\mathbf{XX}} \hat{\mathbf{v}}_{1,\kappa-1} - r_\kappa \widetilde{\mathbf{\Phi}}_{\mathbf{NN}} \hat{\mathbf{v}}_{1,\kappa-1} \right) \\ \mu_\kappa^{(\xi)} \left(\hat{\mathbf{\Phi}}_{\mathbf{XX}} \hat{\mathbf{v}}_{1,\kappa-1} - \xi_\kappa \widetilde{\mathbf{\Phi}}_{\mathbf{NN}} \hat{\mathbf{v}}_{1,\kappa-1} \right) \end{cases} \tag{E.16}$$

mit dem Rayleigh Quotienten

$$r_\kappa = \frac{\hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{\Phi}}_{\mathbf{XX}} \hat{\mathbf{v}}_{1,\kappa-1}}{\hat{\mathbf{v}}_{1,\kappa-1}^H \widetilde{\mathbf{\Phi}}_{\mathbf{NN}} \hat{\mathbf{v}}_{1,\kappa-1}}, \tag{E.17}$$

der Zielfunktion nach der original Herleitung Gl. (5.66)

$$\xi_\kappa = \Re \left\{ \frac{\hat{\mathbf{v}}_{1,\kappa-1}^H \hat{\mathbf{\Phi}}_{\mathbf{XX}} \widetilde{\mathbf{\Phi}}_{\mathbf{NN}} \hat{\mathbf{v}}_{1,\kappa-1}}{\hat{\mathbf{v}}_{1,\kappa-1}^H \widetilde{\mathbf{\Phi}}_{\mathbf{NN}} \widetilde{\mathbf{\Phi}}_{\mathbf{NN}} \hat{\mathbf{v}}_{1,\kappa-1}} \right\} \tag{E.18}$$

und den Schrittweiten

$$\mu_\kappa^{(r)} = \frac{\rho^{(r)}}{r_\kappa}, \qquad \mu_\kappa^{(\xi)} = \frac{\rho^{(\xi)}}{r_\kappa}. \tag{E.19}$$

Es ergeben sich somit die beiden maximalen Faktoren $\rho_{\max}^{(r)}$ und $\rho_{\max}^{(\xi)}$. Exemplarische Simulationsergebnisse sind in Bild E.2 dargestellt für zwei Nachhallzeiten, $T_{60} = 0{,}05\,\text{s}$ und $T_{60} = 0{,}5\,\text{s}$. Zu sehen sind in der oberen Zeile in (a) und (b) die größten Eigenwerte $\lambda_{N,\max}$ und die kleinsten Eigenwerte $\lambda_{N,\min}$ von $\widetilde{\mathbf{\Phi}}_{\mathbf{NN}}$. In der mittleren Zeile in (c) und (d) ist der maximale Schrittweitefaktor $\rho_{\max}^{(r)}$ für die Version von Gl. (E.16) mit dem Rayleigh Quotienten und in der letzten Zeile in (e) und (f) ist entsprechend der maximale Schrittweitefaktor $\rho_{\max}^{(\xi)}$ für die Version mit der Zielfunktion nach der originalen Herleitung abgebildet. Alle Verläufe sind aufgetragen über der diskreten Frequenz $\Omega_k/(2\pi T)$ für $k = 0, \ldots, 128$ mit $1/T = 12\,\text{kHz}$.

Die in Bild E.2 dargestellten Ergebnisse sowie alle weiteren gemachten Experimente führen zu dem Schluss, dass die Schrittweitefaktoren $\rho^{(r)}, \rho^{(\xi)} < 1$ gewählt werden sollten um Stabilität zu gewährleisten.

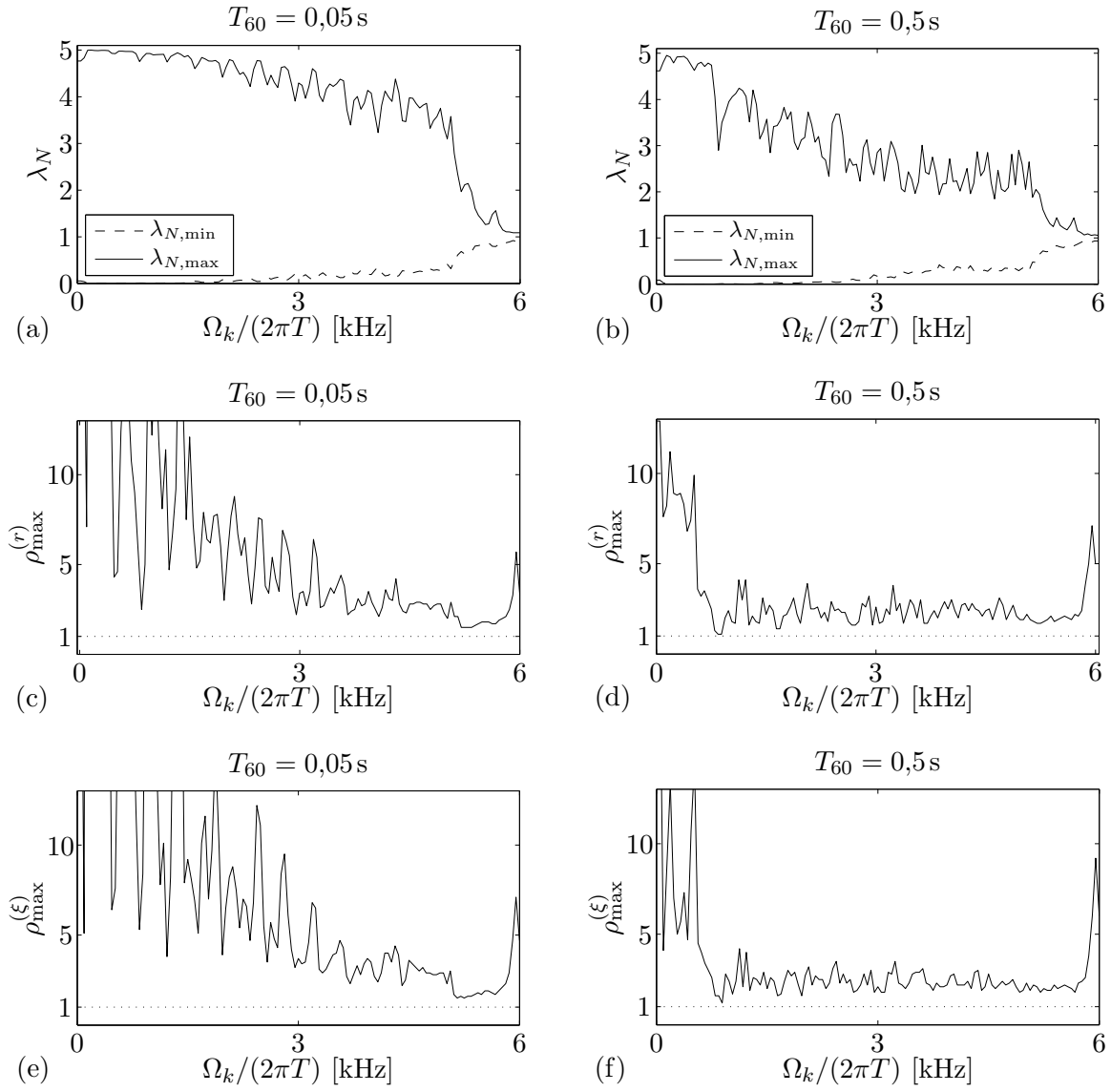

Bild E.2: In (a) und (b) der Verlauf des größten und kleinsten Eigenwertes von $\widetilde{\mathbf{\Phi}}_{\mathbf{NN}}$. Maximaler Schrittweitefaktor für die Version von Gl. (E.16) mit dem Rayleigh Quotienten als Zielfunktion in (c) und (d) sowie für die Version mit der Zielfunktion nach der originalen Herleitung in (e) und (f).

Anhang F

Exkurs zur blinden Quellentrennung

Im Folgenden soll ein Mehr-Sprecher-Szenario mit P Quellen und M Mikrophonen betrachtet werden, wobei $M \geq P$ gilt. Das Signal der i-ten Quelle im Frequenzbereich sei mit $Q_i(\Omega)$ beschrieben, wodurch sich der Vektor für alle Quellen als $\mathbf{Q}(\Omega) = (Q_1(\Omega), .., Q_P(\Omega))^T$ schreiben lässt. Entsprechend existieren P Raumübertragungsfunktionsvektoren $\mathbf{H}_i(\Omega)$, $i = 1, \ldots, P$ zwischen den Quellen und den Mikrophonen, die die so genannte Mischungsmatrix bilden

$$\mathbf{H}(\Omega) = \begin{bmatrix} H_{1,1}(\Omega) & H_{2,1}(\Omega) & \ldots & H_{P,1}(\Omega) \\ H_{1,2}(\Omega) & \ddots & & \vdots \\ \vdots & & & \\ H_{1,M}(\Omega) & \ldots & & H_{P,M}(\Omega) \end{bmatrix} \tag{F.1}$$

$$= [\mathbf{H}_1(\Omega), \mathbf{H}_2(\Omega), .., \mathbf{H}_P(\Omega)]. \tag{F.2}$$

Für das mehrkanalige Mikrophonsignal ergibt sich dann

$$\mathbf{X}(\Omega) = \sum_{i=1}^{P} \mathbf{H}_i(\Omega) Q_i(\Omega) + \mathbf{N}(\Omega) \tag{F.3}$$

$$= \mathbf{H}(\Omega)\mathbf{Q}(\Omega) + \mathbf{N}(\Omega), \tag{F.4}$$

wobei $\mathbf{N}(\Omega) = (N_1(\Omega), \ldots, N_M(\Omega))^T$ einen M-kanaligen Rauschterm beschreibt (die einzelnen Pfade i seien unkorreliert zueinander). Das Ziel der akustischen Quellentrennung besteht nun darin, ein System zu entwickeln, welches aus dem Gemisch der Sprachsignale an den Mikrophonen alle Quellsignale extrahiert. Dieses kann allgemein als MIMO-System (*Multiple-Input Multiple-Output*) bezeichnet werden. Soll die Realisierung ohne Informationen über die *Array*-Geometrie und die Quellenposition erfolgen, so wird sie in der Regel auch als blinde Quellentrennung (engl. *Blind Source Separation*, BSS) bezeichnet. Eine besondere Problemstellung ist hierbei die Tatsache, dass alle Quellen gleichzeitig aktiv sein können.

Ein Großteil der Arbeiten zur blinden Quellentrennung in den letzten Jahren basiert darauf, die Eingangsdaten mit Hilfe der *Independent Component Analysis* (ICA) so zu transformieren, dass die Ergebnisse statistisch unabhängig voneinaner sind [HKO01]. Dabei werden Statistiken höherer Ordnung und nichtlineare Kostenfunktionen eingesetzt, wodurch der Rechenaufwand üblicherweise sehr hoch ist. Da die ICA-Ansätze prinzipiell mit instantanen Mischungen arbeiten, wird die Entmischung im Frequenzbereich pro Frequenzkomponente

separat durchgeführt [SMM05]. Dabei entsteht das so genante Permutationsproblem, d. h. die Zuordnung der separierten frequenzabhängigen Daten zu den entsprechenden Quellen ist nicht eindeutig. Die Zuordnung aller entmischten Frequenzkomponenten jeweis zu den zugehörigen Quellen muss noch mit weiteren Algorithmen explizit durchgeführt werden. Ein Ansatz hierbei ist, adaptive *Beamformer* mit geometrischen Nebenbedingungen und die Verfahren zur BSS zu kombinieren [PA02, KAM07]. Dabei ist jedoch anzumerken, dass solche Methoden nicht mehr blind arbeiten.

Grundsätzlich ist vom physikalischen Standpunkt her die Separation von zwei akustischen Quellen durch BSS-Verfahren im Frequenzbereich äquivalent zum so genannten Null-*Beamforming* mittels zweier adaptiver *Beamformer*. In beiden Fällen wird das Signal der störenden Quelle gedämpft, indem ein Minumum an der korrespondierenden Stelle der Richtcharakteristik der Filterkoeffizienten geformt wird, welche zu der anderen, der gewünschten Quelle gehören [SMH$^+$03]. Dabei ist die Leistungsfähigkeit der BSS-Verfahren limitiert durch die Leistungsfähigkeit von perfekt adaptierten *Beamformern* [Mak03]. Diese haben allerdings den Vorteil, dass die separierten Signale unverzerrt bleiben, unter der Voraussetzung, die jeweiligen Sprecherpositionen zu kennen. Diese sind jedoch gerade in einer verhallten Umgebung bei gleichzeitiger Aktivität der Quellen sehr schwierig zu bestimmen.

In diesem Kapitel soll gezeigt werden, wie mit Hilfe blinder PCA *Beamformer* ein mehrkanaliges Gemisch von zwei Quellsignalen separiert werden kann. Dabei bleiben die Vorteile der räumlichen Filterung erhalten: trotz der Adaption im Frequenzbereich entsteht kein Permutationsproblem, und die Ausgangssignale sind nur geringfügig verzerrt.

F.1 Unterbesetzter Zeit-Frequenz-Raum

Obwohl für die Herleitung des statistisch optimalen *Beamformings* von stationären Signalen ausgegangen wurde, sind Sprachsignale an sich instationäre Zufallssignale. Denn gerade in der zeitlichen Änderung der statistischen Eigenschaften liegt die Information der gesprochenen Sprache. Betrachtet man also das Spektrum einer Äußerung über der Zeit, so kann die spektrale Zusammensetzung erheblich schwanken. Weiterhin kann im Allgemeinen eine deutliche Unterbesetzung der Zeit-Frequenz-Darstellung beobachtet werden (engl. *Time-Frequency-Sparseness*): nur wenige Spektralkomponenten tragen pro betrachteten Zeitabschnitt einen Großteil der Energie. Dabei ist insbesondere die grobe Klassifikation in stimmhafte und stimmlose Sequenzen sehr aufschlussreich. Bei den stimmhaften Lauten konzentriert sich die Energie auf die Stimmbandgrundfrequenz und ihre harmonischen Oberschwingungen. Stimmlose, rauschähnliche Laute weisen ein gleichmäßigeres Spektrum im oberen Spektralbereich auf. Diese Energieverteilung und die Unterbesetzung im Zeit-Frequenz-Raum kann mit Hilfe des Korrelationskoeffizienten zwischen zwei Signalen in unterschiedlichen Frequenzen dargestellt werden. Die Synchronität der Amplituden von verschiedenen Frequenzen soll hier beispielhaft nach [AK00, Ane01] durch die Amplitudenmodulationskorrelation (engl. *Amplituden Modulation Correlation*, AMCor) veranschaulicht werden. In der normierten Form soll der Korrelationskoeffizient der AMCor zwischen zwei Signalen im Frequenzbereich $Q_i(\Omega_k)$ und $Q_j(\Omega_l)$ für die k-te bzw. l-te Frequenzkomponente definiert sein zu

$$\rho(Q_i(\Omega_k), Q_j(\Omega_l)) = \frac{c(Q_i(\Omega_k), Q_j(\Omega_l))}{\sqrt{c(Q_i(\Omega_k), Q_i(\Omega_k)) \cdot c(Q_j(\Omega_l), Q_j(\Omega_l))}} \tag{F.5}$$

mit

$$c(Q_i(\Omega_k), Q_j(\Omega_l)) = E\{|Q_i(\Omega_k)||Q_j(\Omega_l)|\} - E\{|Q_i(\Omega_k)|\}E\{|Q_j(\Omega_l)|\}. \tag{F.6}$$

Im Folgenden soll beispielhaft der Autokorrelationskoeffizient $\rho(Q_1(\Omega_k), Q_1(\Omega_l))$ und der Kreuzkorrelationskoeffizient $\rho(Q_1(\Omega_k), Q_2(\Omega_l))$ ausgewertet werden. Der Erwartungswert in Gl. (F.6) wird über eine zeitliche, blockweise Mittelung realisiert, wobei die Blöcke mit einem Hamming-Fenster der Länge 64 ms und einem Überlapp von 50 % den zu analysierenden Signalen entnommen wurden. Der Betrag der frequenzabhängigen Auto- und Kreuzkorrelationskoeffizienten ist in Bild F.1 in Form einer zweidimensionalen Darstellung von Grauwerten abgebildet. Große Werte für den Betrag des Korrelationskoeffizienten werden durch dunkle Graustufen und kleine Werte durch helle Graustufen charakterisiert.

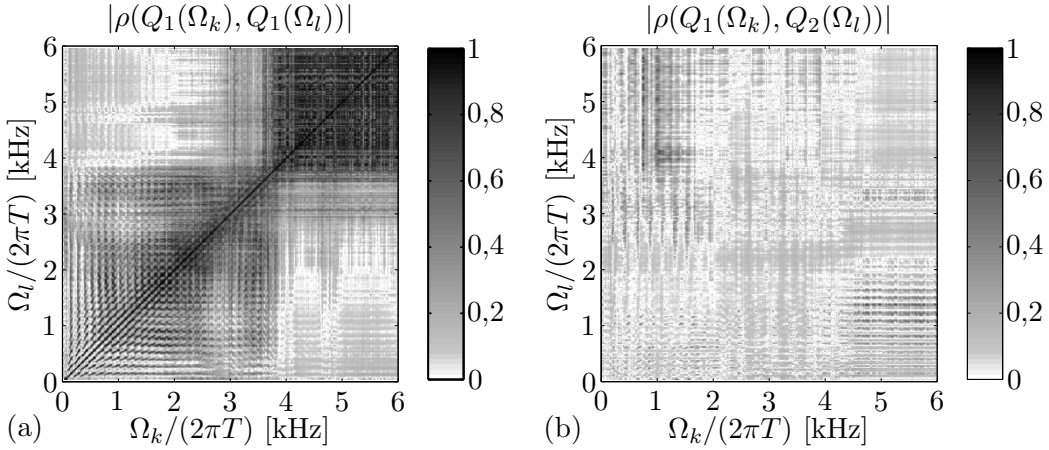

Bild F.1: Betrag des Korrelationskoeffizienten der AMCor für ein Signal in (a) und zwei verschiedene Signale in (b).

In Bild F.1 (a) kann an dem Autokorrelationskoeffizienten bis etwa 4 kHz an dem regelmäßigen Muster die Korrelation der harmonischen Oberschwingungen der Stimmbandgrundfrequenz erkannt werden. Die Abstände der jeweiligen Maxima hängen von der Stimmbandgrundfrequenz ab. Weiterhin sind hohe Korrelationswerte bei benachbarten Frequenzomponenten zu beobachten, die sich an den ausgeprägten Maxima in der Nähe der Diagonalen bemerkbar machen. Bei höheren Frequenzen ab ca. 4 kHz ist ein relativ homogener Bereich zu erkennen, der auf der gleichmäßigeren Amplitudenverteilung der stimmlosen Laute beruht.

Das Bild F.1 (b) zeigt den Kreuzkorrelationskoeffizienten zweier Sprachsignale unterschiedlicher Sprecher. An den hellen Graustufen sind die geringen Werte für die Korrelation der Amplitudenwerte zu erkennen. Die Synchronität der Amplitudenmodulation der analysierten Signale ist somit sehr gering. Dies gilt ebenfalls für den Fall unterschiedlicher Äußerungen desselben Sprechers [AK00]. In [Ane01] wurden erfolgreich Verfahren zur blinden Quellentrennung entwickelt, welche auf Methoden der Dekorrelation mit Hilfe der Amplitudenmodulation basieren.

Basierend auf einem ähnlichen Ansatz der Unterbesetzung im Zeit-Frequenz-Raum sind in den Arbeiten [JRY00, RBR01, YR04] einkanalige Verfahren zur Quellentrennung eingesetzt worden. Dabei berechnet man nun nicht mehr den Grad der Korrelation eines Amplitudenpaars wie in Gl. (F.5) sondern geht per se von einer so genannten disjunkten Orthogonalität[1]

[1] In [JRY00] wird die disjunkte Orthogonalität in einer etwas allgemeineren Form verwendet, da von Signalen ausgegangen wird, welche mit einer Funktion $W(t)$ im Zeitbereich gefenstert wurden, und somit auch der Begriff W-Disjoint Orthogonality gerechtfertigt ist.

(engl. *Disjoint Orthogonality*) aus. Hierbei wird bei der blockweisen Verarbeitung für jedes Signalpaar i, j die disjunkte Orthogonalität pro Block m zu $Q_{i,m}(\Omega_k) \cdot Q^*_{j,m}(\Omega_k) = 0$, $\forall m, k$ mit $i \neq j$ definiert[2]. Mit der realistischen Annahme einer approximativen disjunkten Orthogonalität

$$Q_{i,m}(\Omega_k) \cdot Q^*_{j,m}(\Omega_k) \approx 0, \qquad \forall m, k \quad i \neq j \tag{F.7}$$

gelangt man zu der Idee, pro Zeit-Frequenz-Punkt ein einkanaliges Signal "an- und abzuschaltet" bei der jeweiligen Dominanz einer bestimmten Quelle. Eine solche binäre Maskierung (engl. *Binary Masking*, BM) wird im DUET-Algorithmus (*Degenerate Unmixing Estimation Technique*) vorgenommen [JRY00, RBR01, YR04], wobei die Dominanz mit Hilfe von Amplituden- und Phaseninformationen eines mehrkanaligen Signals bestimmt wird. Die binäre Maske soll wie folgt definiert sein

$$\varpi^{(\mathrm{BM})}_{i,m}(\Omega_k) = \begin{cases} 1, & \text{für} \quad |Q_{i,m}(\Omega_k)| > v_g \cdot |Q_{j,m}(\Omega_k)|, \qquad \forall i \neq j \\ 0, & \text{sonst} \end{cases} \tag{F.8}$$

wobei $v_g \in \mathbb{R}^+$ ein heuristischer Parameter ist. Die entmischten Signale ergeben sich dann zu

$$\hat{Q}_{i,m}(\Omega_k) = \varpi^{(\mathrm{BM})}_{i,m}(\Omega_k) X_{1,m}(\Omega_k). \tag{F.9}$$

Auch wenn die binäre Maske Gl. (F.8) jeweils optimal bestimmt wird, kann die Qualität der entmischten Signale durch das harte An- und Abschalten erheblich schwanken [WHUTV07].

F.2 PCA Beamforming im Mehr-Sprecher-Szenario

Motiviert durch die starke Unterbesetzung des Zeit-Frequenz-Raums soll das PCA *Beamforming* zur blinden Quellentrennung akustischer Signale eingesetzt werden. Die Idee hierbei liegt darin, mehrere PCA *Beamformer* zu verwenden, und für jeden Zeit-Frequenz-Punkt diejenige frequenzabhängige PCA-Adaptionsregel zu aktivieren, welche der entsprechenden dominanten Quelle zugewiesen wurde. Dafür wird in der Adaptionsregel Gl. (5.37) im additiven Term der Koeffizientenänderung die binäre Maskierung hinzugefügt. Dadurch erfolgt eine Änderung der Filterkoeffizienten $\mathbf{F}_{i,m}(\Omega_k)$ des i-ten PCA *Beamformers* nur dann, wenn die zugehörige Quelle $Q_{i,m}(\Omega_k)$ für diesen Zeitpunkt m und diese Frequenzkomponente Ω_k dominant ist. Die Adaptionsregel lautet folglich

$$\begin{aligned} \mathbf{F}_{i,m}(\Omega_k) = {} & \frac{M^{-1} + \mathbf{F}^H_{i,m-1}(\Omega_k)\mathbf{F}_{i,m-1}(\Omega_k)}{2\mathbf{F}^H_{i,m-1}(\Omega_k)\mathbf{F}_{i,m-1}(\Omega_k)} \mathbf{F}_{i,m-1}(\Omega_k) \\ & + \varpi^{(\mathrm{BM})}_{i,m}(\Omega_k)\mu_{i,m}(\Omega_k) Y^*_{i,m}(\Omega_k) \left(\mathbf{X}_m(\Omega_k) - \frac{Y_{i,m}(\Omega_k)}{\mathbf{F}^H_{i,m-1}(\Omega_k)\mathbf{F}_{i,m-1}(\Omega_k)} \mathbf{F}_{i,m-1}(\Omega_k) \right). \end{aligned} \tag{F.10}$$

In Gl. (F.10) ist der *Constraint* nach Abschnitt 6.4.1 zu $C^2 = M^{-1}$ gesetzt. Die Schrittweite des i-ten *Beamformers* $\mu_{i,m}(\Omega_k)$ soll abhängig sein von der Frequenz und der Zeit, und der Ausgang ergibt sich zu $Y_{i,m}(\Omega_k) = \mathbf{F}^H_{i,m-1}(\Omega_k)\mathbf{X}_m(\Omega_k)$. Das Adaptionsschema ist hier das

[2] Zu beachten ist der Unterschied zur statistischen Orthogonalität $E\{Q_i(\Omega_k) \cdot Q^*_j(\Omega_k)\} = 0$, welche über alle Realisierungen von $Q_i(\Omega_k)$ und $Q_j(\Omega_k)$ entsteht.

gleiche wie bei Algorithmus 4 (S-Grad-IS), jedoch muss dann der Faktor $\varpi_{i,m}^{(\mathrm{BM})}(\Omega_k)$ für die Adaptionssteuerung eingefügt werden.

Nach *Szenario-5* wurden mehrkanalige Mischsignale für den Fall von $P = 2$ Quellen erzeugt, wobei die Leistungen der Signale beider Quellen gleich groß sind. Die beiden Quellen befinden sich jeweils im Abstand von 2 m zum *Array* mit der Richtung von $\theta_{s1} = 45^\circ$ für die eine und $\theta_{s2} = \text{-}30^\circ$ für die andere Quelle relativ zur *Broadside*-Ausrichtung. Zur Entmischung sind daher zwei PCA *Beamformer* notwendig, die jeweils mit 256 Koeffizienten pro Filter und einer jeweiligen Länge von 512 für die Fourier-Transformation realisiert wurden. Da bei Simulationen die Quellsignale bekannt sind, kann eine optimale binäre Maske berechnet werden. Dabei ist der Grenzwert v_g in Gl. (F.8) für die Dominanz einer Quelle für die Frequenzkomponente Ω_k und den Block m so gewählt, dass eine Quelle als Dominat gilt, wenn deren Leistung mindestens 6 dB größer als die Leistung der jeweils anderen Quelle ist. Die sich so ergebenden Richtdiagramme der beiden PCA *Beamformer* sind beispielhaft für den Fall einer Freifeldanordnung in Bild F.2 dargestellt. Bei der Adaption waren beide Quellen simultan aktiv, wobei die Werte $\varpi_{i,m}^{(\mathrm{BM})}(\Omega_k)$ in Gl. (F.10) optimal bestimmt wurden.

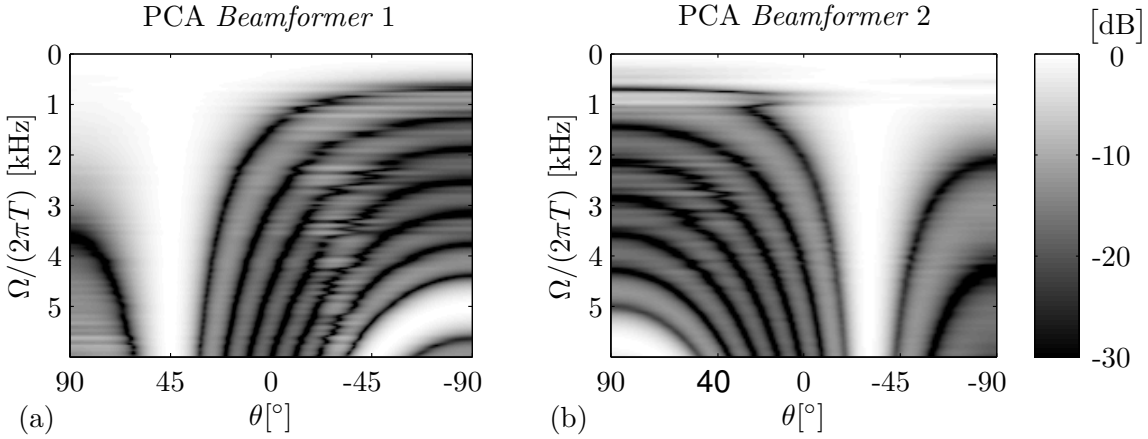

Bild F.2: Richtdiagramme der beiden PCA *Beamformer* bei zwei aktiven Quellen mit den Richtungen $\theta_{s1} = 45^\circ$ und $\theta_{s2} = \text{-}30^\circ$ und der Verwendung von $M = 8$ Sensoren. Die binären Masken zur Adaptionssteuerung wurden optimal bestimmt.

Die Ergebnisse in Bild F.2 veranschaulichen, dass sich die jeweiligen *Beamformer* auf die beiden Quellen ausgerichtet haben, obwohl beide Sprecher gleichzeitig aktiv waren. Voraussetzung hierbei ist natürlich, dass die Werte $\varpi_{i,m}^{(\mathrm{BM})}(\Omega_k)$ korrekt ermittelt werden. Denn nur wenn tatsächlich die Dominanz einer Quelle vorherrscht, kann der zugehörige *Beamformer* und dessen Filterkoeffizienten für diesen Zeit-Frequenz-Punkt einen Adaptionsschritt "in die richtige Richtung" machen. Aus dieser Sicht heraus kann die Adaptionssteuerung auch etwas allgemeiner formuliert werden. Wenn die Wahrscheinlichkeit hoch ist, dass eine bestimmte Quelle für einen Zeit-Frequenz-Punkt dominant ist, sollten sich die Filterkoeffizenten stärker ändern können als für den Fall, dass die Wahrscheinlichkeit für die Dominanz gering ist. Oder anders ausgedrückt, je dominanter eine bestimmte Quelle für einen Zeit-Frequenz-Punkt ist, desto stärker sollten sich die Filterkoeffizenten ändern können. Mit dieser Erkenntnis soll eine wahrscheinlichkeitsbasierte Maskierung (engl. *Likelihood Masking*, LM) vorgeschlagen werden:

$$\varpi_{i,m}^{(\mathrm{LM})}(\Omega_k) \approx p(|Q_{i,m}(\Omega_k)| \gg |Q_{j,m}(\Omega_k)| \,|\, \mathbf{X}_m(\Omega_k)), \quad \forall j,\ j \neq i. \tag{F.11}$$

In Gl. (F.11) bezeichnet also $p(|Q_{i,m}(\Omega_k)| \gg |Q_{j,m}(\Omega_k)| \,|\, \mathbf{X}_m(\Omega_k))$ die Wahrscheinlichkeit da-

für, dass die i-te Quelle für die k-te Spektralkomponente und den m-ten Verarbeitungsblock, gegeben die mehrkanaligen Eingangsdaten wesentlich dominanter als alle anderen Quellen ist. In [WHUTV07] wurde ein Verfahren vorgestellt, welches mittels Dekorrelationsfiltern, jeweils angeordnet zwischen benachbarten Mikrophonen eine grobe Vorseparation der Quellsignale vornimmt, die dann ins Verhältnis gesetzt einen Wert für die *Likelihood*-Maskierung liefern. Diese Methode setzt ein äquidistantes, lineares Mikrophon-*Aarray* voraus, weshalb die Bezeichnung symmetrisch adaptive Dekorrelation (engl. *Symmetric Adaptive Decorrelation*, SAD) eingeführt wurde. Andere Methoden zur Bestimmung von $\varpi_{i,m}^{(\mathrm{LM})}(\Omega_k)$, wie z. B. die Ausnutzung von Phasen- und Dämpfungseigenschaften der zeitversetzten Mikrophonsignale wie in [RBR01] oder die Auswertung der Amplitudenmodulation wie in [AK00] sind aktueller Forschungsgegenstand. An dieser Stelle soll lediglich die Möglichkeit der Separation von akustischen Signalen mittels PCA *Beamforming*, gegeben eine perfekte binäre Maskierung, demonstriert werden.

Nun, da mit Hilfe der zusätzlichen Adaptionssteuerung eine Adaption der PCA *Beamformer* hin zu den verschiedenen Quellen möglich ist, soll noch eine Weiterverarbeitung der Filterkoeffizienten erfolgen. Denn, obschon das Maximum des *Beampatterns auf den Zielsprecher ausgerichtet ist, erfolgt keine* explizite Minima-Bildung an den Stellen der anderen Quellen, wie an den Richtdiagrammen in Bild F.2 zu erkennen ist. Dies ist in der PCA-Adaptionsregel ja auch nicht vorgesehen. Daher soll eine gegenseitige orthogonale Projektion (engl. *Mutual Orthogonal Projection*, MOP) den PCA-Filterkoeffizienten nachgeschaltet werden. Dazu wird für jede Frequenzkomponente aus dem System linear unabhängiger Filtervektoren der Quellen $j \neq i$ ein orthogonaler Untervektorraum erzeugt, in den der Filtervektor der Quelle i hineinprojeziert wird:

$$\mathbf{W}_{i,m}(\Omega_k) = \left(\prod_{j;j\neq i} \left[\mathbf{I} - \mathbf{F}_{j,m}(\Omega_k) \mathbf{F}_{j,m}^{H}(\Omega_k) \right] \right) \mathbf{F}_{i,m}(\Omega_k). \tag{F.12}$$

Verwendet man die Filterkoeffizienten der beiden PCA *Beamformer*, welche die Richtdiagramme in Bild F.2 erzeugen, in der orthogonalen Projektion Gl. (F.12), so führen die resultierenden Koeffizienten zu den *Beampattern* in Bild F.3. Dort sind nun neben den Maxima für die Richtungen der Zielquellen auch Minima zu beobachten, jeweils an der Stelle der anderen Quelle.

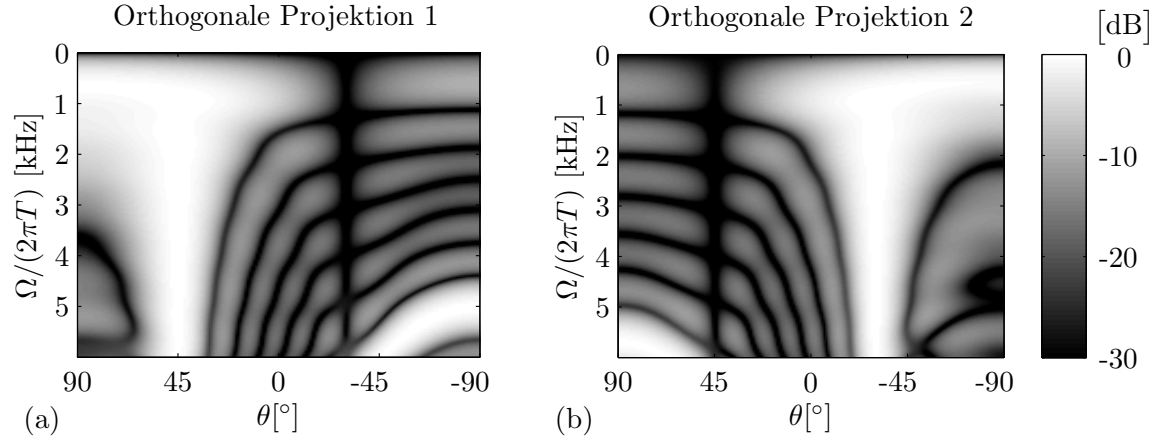

Bild F.3: Richtdiagramme der beiden *Beamformer* nach der orthogonalen Projektion für zwei aktive Quellen mit den Richtungen $\theta_{s1} = 45\,^\circ$ und $\theta_{s2} = -30\,^\circ$.

Zur Filterung der breitbandigen Sprachsignale sollen nicht die durch die Projektion bestimmten Filterkoeffizienten in Gl. (F.12) direkt verwendet werden. Sondern es soll die dominante Komponente in dem orthogonalen Unterraum explizit berechnet werden, welche dann zur Reproduktion der Quellsignale führt. Für die Realisierung in einer adaptiven Gesamtstuktur hat sich dabei der Einsatz einer weiteren PCA-Adaption pro orthogonalem Filtersatz als effektiv erwiesen [TV07]. Die Adaptionsregel für den PCA *Beamformer* mit orthogonaler Nebenbedingung (engl. *Orthogonal Constraint Principal Component Analysis*, OCPCA) ist in Algorithmus 11 (OCPCA) beschrieben. Wie in den anderen Adaptionsalgorithmen soll auch hier auf den frequenzabhängigen Parameter Ω_k verzichtet werden, da das Vorgehen für jede Spektralkomponente gleich ist und so die Darstellung übersichtlicher wird.

Algorithmus 11 (OCPCA) *Wähle die Glättungskonstante α, den Schrittweitefaktor ρ, den Constraint C und die Startvektoren $\mathbf{W}_{i,0} \in \mathbb{C}^M$, $i = 1, \ldots, M$. Berechne bei gegebenen PCA-Filterkoeffizienten $\mathbf{F}_{j,m}$, $j = 1, \ldots, M$ für alle Verarbeitungsblöcke $m = 1, 2, \ldots$ und für alle OCPCA Beamformer $i = 1, \ldots, M$*

$$
\begin{aligned}
\hat{Q}_{i,m} &:= \mathbf{W}_{i,m-1}^H \mathbf{X}_m \\
\tilde{\mu}_{i,m}^{-1} &:= \alpha \tilde{\mu}_{i,m}^{-1} + (1-\alpha)|\hat{Q}_{i,m}|^2 \\
\mu_i &:= \tilde{\mu}_{i,m} \varrho C^2 \\
\mathbf{V}_i &:= \mathbf{W}_{i,m-1} + \mu_i \hat{Q}_{i,m}^* \left(\mathbf{X}_m - \hat{Q}_{i,m} \mathbf{W}_{i,m-1} \right) \\
\tilde{\mathbf{V}}_i &:= \left(\prod_{j;j\neq i} [\mathbf{I} - \mathbf{F}_{j,m} \mathbf{F}_{j,m}^H] \right) \mathbf{V}_i \\
\tilde{\mathbf{W}}_i &:= \frac{\tilde{\mathbf{V}}_i}{C \tilde{V}_{1,i}} \\
R_i^2 &:= \tilde{\mathbf{W}}_i^H \tilde{\mathbf{W}}_i \\
\mathbf{W}_{i,m} &:= \frac{C^2 + R_i^2}{2R_i^2} \tilde{\mathbf{W}}_i .
\end{aligned}
$$

Anmerkungen zum Algorithmus 11 (OCPCA) Bei der Filterung $\mathbf{W}_{i,m-1}^H \mathbf{X}_m$ zur Schätzung der Quellsignale ist auf zyklische Effekte zu achten. Dies kann effizient durch das Overlap-Save-Verfahren geschehen. Weiterhin erfolgt die Subtraktion $\mathbf{X}_m - \hat{Q}_{i,m} \mathbf{W}_{i,m-1}$ im Zeitbereich. Die Normierung direkt nach der orthogonalen Projektion mit dem ersten Element des Vektors $\tilde{\mathbf{V}}_i = (\tilde{V}_{1,i}, \ldots, \tilde{V}_{M,i})^T$ hat sich bei den Experimenten als deutliche Robustheitssteigerung erwiesen. Diese Normierung kann auch mittels adaptiver Methoden recheneffizienter durchgeführt werden [TV07]. Die Norm der Filterkoeffizienten wird mit der Division von $\tilde{\mathbf{V}}_i$ durch C und der abschließenden Newton-Iteration auf den Wert C festgelegt. Wichtig ist hier noch anzumerken, dass im Gegensatz zu Gl. (F.10) die OCPCA-Filterkoeffizienten permanent adaptiert werden können.

Zur Beurteilung der Separationsleistung sollen für die folgenden Simulationen konvergierte Filterkoeffizienten für die beiden PCA und OCPCA *Beamformer* angenommen werden. Besteht das Eingangssignal nun nur aus dem ersten Quellsignal, so sollte dieses an dem ersten OCPCA-Ausgang möglichst unverzerrt beobachtet werden und entsprechend an dem zweiten OCPCA-Ausgang komplett unterdrückt sein. Bei der Filterung des zweiten Quellsignals an der anderen räumlichen Position sollte sich das Verhalten umkehren: das Signal liegt am zweiten OCPCA-Ausgang vor. Dadurch lässt sich pro Ausgang das Verhältnis der Leistungen

des gewünschten Zielsignals zum störenden Quellsignal (engl. *Signal-to-Interference-Ratio*, SIR) bestimmen. Die Sprachqualität der ermittelten Zielsignale kann relativ zu dem verhallten, reinen Sprachsignal an einem Mikrophon bewertet werden. Bei den Simulationen nach *Szenario-5* ergeben sich dadurch pro untersuchter Nachhallzeit 90 PSM-Werte[3]. Um deutlich zu machen, dass die Separationsleistung für die beiden Ausgänge unterschiedlich sein kann, wurden für jede Nachhallzeit und jede Quellenkombination die höheren PSM-Werte und die niedrigeren PSM-Werte gesondert gemittelt. Diese sind in Bild F.4 in der linken Spalte für die Anordnungen bestehend aus $M = 5$ und $M = 9$ Mikrophonen dargestellt und mit "hoch" für die größeren PSM-Werte, sowie mit "niedrig" für die kleineren PSM-Werte bezeichnet. Außerdem ist noch der Mittelwert aller Werte aufgetragen ("mittel"). Das gemittelte SIR für diese beiden Gruppen ist in der rechten Spalte von Bild F.4 zu sehen. Dabei bezeichnet "SIR PSM-hoch" die gemittelten SIR-Werte aus der Gruppe der Sprachbeispiele mit den höheren PSM-Werten und entsprechend "SIR PSM-gering" das gemittelte SIR für die Gruppe mit den kleineren PSM-Werten. Zusätzlich ist der gesamte Mittelwert dargestellt ("SIR mittel").

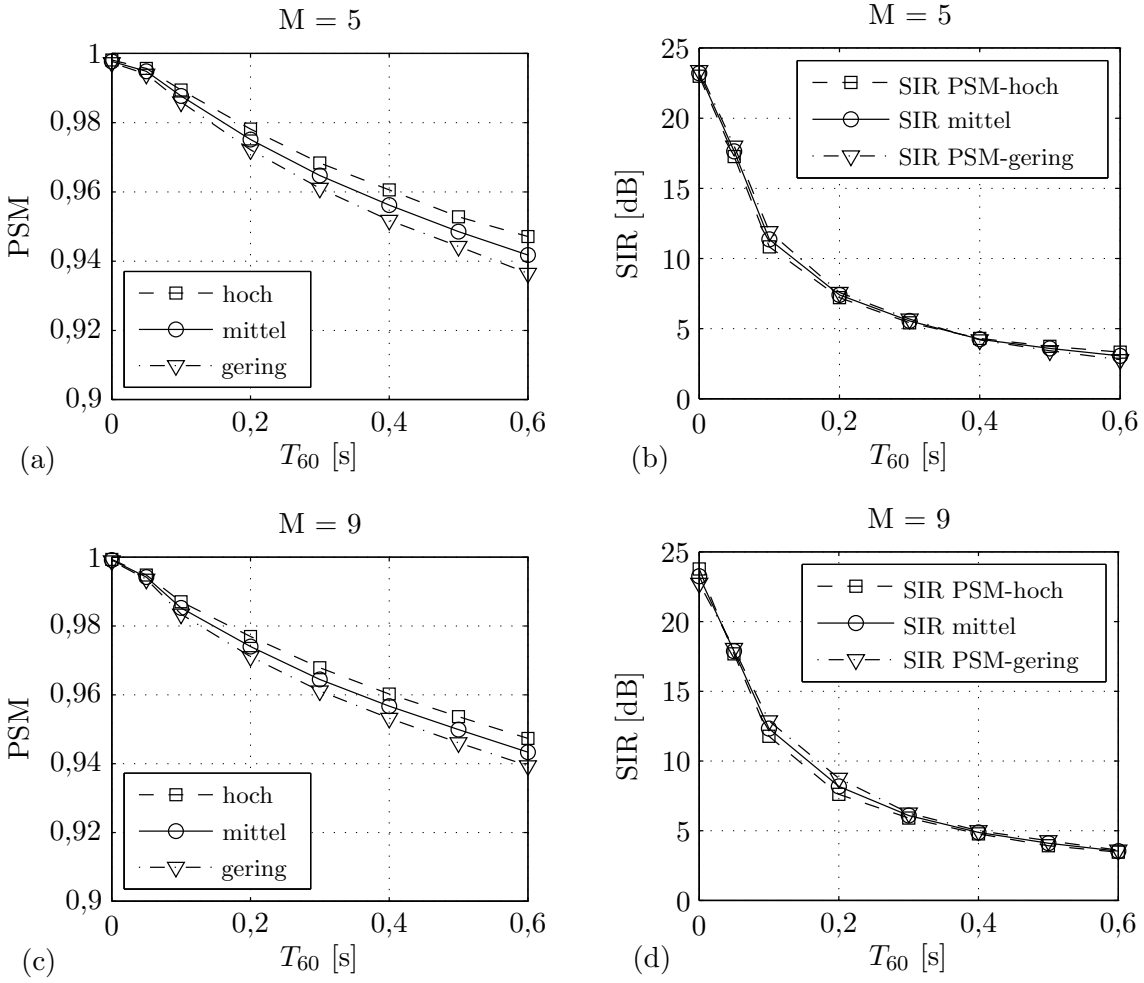

Bild F.4: Perzeptives Qualitätsmaß und SIR für die Quellentrennung nach der Anordnung *Szenario-5* für $M = 5$ und $M = 9$ Mikrophone. Alle Filterkoeffizienten sind im konvergierten Zustand und haben jeweils eine Länge von 256.

Wie bei allen Verfahren zur akustischen blinden Quellentrennung nimmt die Separations-

[3]Bei der Auswahl von 2 aus 10 verschiedenen Sprachsignalen ergeben sich 45 unterschiedliche Kombinationen. Nach der Verarbeitung liegen somit 90 Schätzungen für die Quellsignale vor.

leistung mit steigender Nachhallzeit deutlich ab. An den Verläufen des perzeptiven Qualitätsmaßes ist zwar ein Unterschied zwischen den Ausgängen mit den besseren und den schlechteren PSM-Werten zu sehen. In [WHUTV07] wurde demonstriert, dass dieser jedoch z. B. im Vergleich zur Quellentrennung mit dem DUET-Verfahren sehr gering ist. An den leicht niedrigeren SIR-Werten in Bild F.4 für die Gruppe der Signale mit der besseren Sprachqualität ist zu erkennen, dass sich die Störsignalunterdrückung prinzipiell konträr zur Sprachqualität verhält. Insgesamt ist zu der Sprachqualität noch anzumerken, dass diese bei subjektiven Hörtests zwar sehr gut ist, aber eine Tendenz zur Dämpfung tiefer Frequenzen bei steigender Nachhallzeit vorhanden ist. Dadurch sind die fallenden PSM-Ergebnisse in Bild F.4 (a) und (c) zu erklären. Weiterhin ist noch interessant, dass die Unterschiede zwischen der Anordnung mit 5 und mit 9 Mikrophonen gering sind. Bei der Verwendung von weniger als 5 Mikrophonen ist die Änderung jedoch größer.

Auch wenn das Maximum der räumlichen Übertragungsfunktion an der Stelle der Zielquelle und das Minimum an der Stelle der zu unterdrückenden Störquelle liegen, sinkt das SIR deutlich bei steigenden Nachhallzeiten aufgrund der Mehrwegeausbreitung. Hier ist eine weitere Steigerung der Störsignalunterdrückung durch die Nachschaltung eines einkanaligen Filters möglich. Solch ein Vorgehen ist vergleichbar mit dem Wiener *Post Filter* bei dem MMSE *Beamformer* in Gl. (4.50). Ein äquivalentes *Post Filter* ist in [WHUTV07] erfolgreich eingesetzt worden (siehe [TV07] für eine genaue Beschreibung dieser Methode).

F.3 Zusammenfassung

In diesem Kapitel wurde die Möglichkeit zur akustischen Quellentrennung mittels blinder PCA *Beamformer* demonstriert. Dabei wird ausgenutzt, dass trotz simultaner Aktivität verschiedener Quellsignale der Zeit-Frequenz-Raum unterbesetzt ist. Gelingt es also, die Dominanz einer Zielquelle für einen Zeit-Frequenz-Punkt festzustellen, so können genau für diese Frequenzkomponente die entsprechenden PCA-Filterkoeffizienten adaptiert werden. Dafür wurde hier eine optimal bestimmte binäre Maskierung verwendet. Die robuste Adaptionssteuerung mit Hilfe einer wahrscheinlichkeitsbasierten *Likelihood*-Maskierung ist weiterhin aktueller Forschungsgegenstand. Desweiteren wurde hier gezeigt, wie mit Hilfe einer orthogonalen Projektion in einem PCA-Adaptionsschema die Störsignale zu unterdrücken sind.

Formelzeichen und Abkürzungen

Allgemeine Notation

- Vektoren werden durch fettgedruckte Buchstaben dargestellt: $\mathbf{x} = [x_1, x_2, \ldots, x_M]^T$.
- Matrizen werden durch fettgedruckte Buchstaben dargestellt: $\mathbf{\Phi_{XX}}$.
- Schätzgrößen werden durch ein Dach gekennzeichnet und sind nicht immer explizit aufgelistet: $\hat{\theta}$.
- Darstellungen im Frequenzbereich werden durch Großbuchstaben gekennzeichnet: X_1.
- Der tiefergestellte Index m steht für blockabhängige Variablen die nicht immer explizit aufgelistet sind: X_m.
- Der tiefergestellte Index κ steht für iterativ bestimmte Größen die nicht immer explizit aufgelistet sind: $\mathbf{v}_{1,\kappa}$.
- Eingeführte Variablen in den Beschreibungen der Adaptionsschemata werden hier nicht aufgelistet.

Römische Formelzeichen

A	Wandfläche
$A_i(\Omega)$	Übertragungsfunktion zwischen Störquelle und dem i-ten Sensor
$a_i(n)$	(zeitdiskrete) Raumimpulsantwort zwischen Störquelle und dem i-ten Sensor
a_{ij}	Zustandsübergang von der Hypothese H_i nach H_j in einem HMM
B	Blockverschiebung
$B(\Omega, \theta)$	*Beampattern*
$B_{\mathrm{DSB}}(\Omega, \theta)$	*Beampattern* des DSBs
$B_{\mathrm{DSB}}^{(\mathrm{LE})}(\Omega)$	*Beampattern* des DSBs mit Lokalisationsfehler
$\mathbf{B}(\Omega)$	*Blocking Matrix*
$\mathbf{B}_{\mathrm{ASC\,opt}}(\Omega)$	optimale *Adaptive Speech Cancellation Blocking Matrix*
$\mathbf{B}_{\mathrm{ASC}}(\Omega)$	*Adaptive Speech Cancellation Blocking Matrix*
$\mathbf{B}_{\mathrm{DO}}(\Omega)$	*Delay Only Blocking Matrix*
$\mathbf{B}_{\mathrm{DOR}}(\Omega)$	*Delay Only Ratio Blocking Matrix*
$\mathbf{B}_{\mathrm{GEV}}(\Omega)$	*Generalized Eigenvector Blocking Matrix*
$\mathbf{B}_{\mathrm{TF}}(\Omega)$	*Transfer Function Blocking Matrix*
$\mathbf{B}_{\mathrm{TFR}}(\Omega)$	*Transfer Function Ratio Blocking Matrix*
BA	*Blocking Ability*

$\boldsymbol{\mathcal{B}}(\Omega)$	Projektionsvektor der *Blocking Matrix*
C	Parameter für die Nebenbedingung der Gradientenregel
c	Schallgeschwindigkeit
C_{50}	Deutlichkeitsmaß
C_{80}	Klarheitsmaß
c_i	Skalierungsfaktor
$c_{il}(\tau)$	Kreuzkorrelation zwischen zwei Signalen $x_i(t)$ und $x_l(t)$
d	Mikrophonabstand zwischen zwei Sensoren
d_{il}	Mikrophonabstand zwischen dem i-ten Sensor und l-ten Sensor
$D(\Omega)$	Direktivität
$\mathbf{d}(\Omega, \mathbf{p})$	*Steering Vector*
$\mathbf{D}_\kappa$	iterativ bestimmte Matrix zur Einhaltung der Nebenbedingung der neuen Gradientenregel
$DI(\Omega)$	Bündelungsmaß
diag$\{\cdot\}$	Diagonalmatrix
$E\{\cdot\}$	Erwartungswert
$E(t)$	Energie eines Schallereignisses
e	Exponentialfunktion
E_0	Anfangsenergie eines Schallereignisses
$e(\cdot)$	Fehlerfunktion
$\bar{e}(\cdot)$	mittlerer Fehler
$E_A(t)$	Energieabfallkurve
E_D	Energiedichte des Direktschallfelds
E_{St}	Energiedichte des stationären Schallfelds
exp$\{\cdot\}$	Exponentialfunktion
$\mathcal{E}_{i-1,1}(\Omega)$	Fehler zwischen dem $(i-1)$-ten Ausgangssignal der TFRBM und dem ersten Mikrophonsignal
f	kontinuierliche Frequenz
f_{Ab}	Abtastfrequenz
f_k	k-te diskrete Frequenz
$f_i(n)$	(zeitdiskrete) i-te *Beamformer*-Filterimpulsantwort
$\tilde{f}_i(n)$	(zeitdiskrete) i-te zeitinverse *Beamformer*-Filterimpulsantwort
$F_i(\Omega)$	i-te *Beamformer*-Übertragungsfunktion
$\mathbf{F}(\Omega)$	allgemeiner Filterkoeffizientenvektor eines *Beamformers*
$\mathbf{F}_{\mathrm{ref}}(\Omega)$	Referenz-Filterkoeffizienten des *Fixed Beamformers*
$\mathbf{F}_{\mathrm{FBF}}(\Omega)$	Filterkoeffizientenvektor des *Fixed Beamformers*
$\mathbf{F}_{\mathrm{Frost}}(\Omega)$	Filterkoeffizientenvektor des Frost *Beamformers*
$\mathbf{F}_{\mathrm{DSB}}(\Omega)$	Filterkoeffizientenvektor des idealen *Delay-and-Sum-Beamformers*
$\mathbf{F}_{\mathrm{GML}}(\Omega)$	Filterkoeffizientenvektor des GML *Beamformers*
$\mathbf{F}_{\mathrm{GMMSE}}(\Omega)$	Filterkoeffizientenvektor des GMMSE *Beamformers*
$\mathbf{F}_{\mathrm{GMV}}(\Omega)$	Filterkoeffizientenvektor des GMV *Beamformers*
$\mathbf{F}_{\mathrm{GMVDR}}(\Omega)$	Filterkoeffizientenvektor des GMVDR *Beamformers*
$\mathbf{F}_{\mathrm{ML}}(\Omega)$	Filterkoeffizientenvektor des ML *Beamformers*
$\mathbf{F}_{\mathrm{MV}}(\Omega)$	Filterkoeffizientenvektor des MV *Beamformers*
$\mathbf{F}_{\mathrm{MVDR}}(\Omega)$	Filterkoeffizientenvektor des MVDR *Beamformers*
$\mathbf{F}_{\mathrm{MMSE}}(\Omega)$	Filterkoeffizientenvektor des MMSE *Beamformers*

$\mathbf{F}_{\mathrm{MF}}(\Omega)$	*Matched-Filter*-Koeffizienten
$\mathbf{F}_{\mathrm{PCA}}(\Omega)$	Filterkoeffizientenvektor des PCA *Beamformers*
$\mathbf{F}_{\mathrm{PCA}\,\nu}(\Omega)$	diskretisierte *a priori* berechnete PCA-Filterkoeffizienten
$\mathbf{F}^{(\mathrm{SNR})}(\Omega)$	Filterkoeffizientenvektor korrespondierend zu einem dominanten Eigenwert
$\mathbf{F}_{(\mathrm{SNR})}(\Omega)$	definierter Filterkoeffizientenvektor welcher das SNR maximiert
$\tilde{\mathbf{F}}_{\mathrm{SNR}}(\Omega)$	Lösungsvektor des verallgemeinerten Eigenwertproblems
$G(\Omega)$	frequenzabhängiger *Array Gain*
$G_i(\Omega)$	i-te Gewichtungsfunktion des GCCs
$G^W(\Omega)$	*White Noise Gain*
$G^W_{\mathrm{DSB}}(\Omega)$	*White Noise Gain* des idealen *Delay-and-Sum-Beamformers*
$G_{\mathrm{SNR}}(\Omega)$	SNR-Gewinn des statistisch optimalen *Beamformers*
$G^W_{\mathrm{SNR}}(\Omega)$	*White Noise Gain* des statistisch optimalen *Beamformers*
$\boldsymbol{\mathcal{G}}(\Omega)$	adaptiver Filterkoeffizientenvektor im GSC
$\boldsymbol{\mathcal{G}}_{\mathrm{opt}}(\Omega)$	optimaler Filterkoeffizientenvektor im GSC
$H_i(\Omega)$	Übertragungsfunktion zwischen Sprecher und dem i-ten Sensor
$H_0(\Omega_k)$	Hypothese einer Sprachpause
$H_1(\Omega_k)$	Hypothese für Sprachaktivität
$h_i(n)$	(zeitdiskrete) i-te Raumimpulsantwort
i	Laufindex
$I_0\{\cdot\}$	modifizierte Besselfunktion nullter Ordnung
$I_1\{\cdot\}$	modifizierte Besselfunktion erster Ordnung
j	Laufindex
$J_{\mathrm{GSC}}(\Omega)$	Kostenfunktion der ANC-Filterkoeffizienten
$J_{\mathrm{MSE}}(\cdot)$	Kostenfunktion des MSE-Ansatzes
$J_{\mathrm{MV}}(\cdot)$	Kostenfunktion des MV-Ansatzes
k	Laufindex
K_N	Abweichung der Varianzschätzung des Rauschens relativ zur Varianz des Rauschens
K_o	obere Schranke für diskrete Spektralkomponenten
K_S	Abweichung der Varianzschätzung des Rauschens relativ zur Varianz der Sprache
K_u	untere Schranke für diskrete Spektralkomponenten
$\mathcal{K}$	Krylov Unterraum
L	Länge der diskreten Fourier-Transformation
l	Laufindex
l_x	Anzahl der Verarbeitungsblöcke des Signals $x(n)$
l_s	Anzahl der Verarbeitungsblöcke des Sprachsignals $s(n)$
$\bar{l}$	mittlere freie Weglänge des Schalls
$\ln(\cdot)$	natürlicher Logarithmus
L_n	Anzahl der Abtastwerte des Rauschsignals (ohne Sprachanteil)
L_{rel}	relativer Schalldruckpegel
L_s	Anzahl der Abtastwerte des Sprachsignals
$\log_{10}(\cdot)$	10-er Logarithmus
$\mathcal{L}(\cdot)$	*Log-Likelihood*-Funktion
M	Anzahl der Mikrophone

m	Blockindex
$M\{\cdot\}$	konfluent hypergeometrische Funktion
N	Obergrenze von Laufvariablen
n	diskreter Zeitindex
N_D	Intervall der maximal möglichen Verschiebungs-Abtastwerte
$n_i(\theta)$	richtungsabhängige Verschiebung am i-ten Mikrophon
$N_i(\Omega)$	Störsignal am i-ten Sensor im Frequenzbereich
$\bar{n}$	mittlere Stoßzahl des Schalls
$n_c(n)$	(zeitdiskretes) räumlich korreliertes Störsignal
$n_i(n)$	(zeitdiskretes) Störsignal am i-ten Sensor
$n_{u,i}(n)$	(zeitdiskretes) räumlich unkorreliertes Störsignal am i-ten Sensor
n_0	Zeitindex für das Maximum der Impulsantwort
n_{50}	Zeitindex korrespondierend zur Zeit 50 ms
$\mathcal{O}\{\cdot\}$	Komplexitätsordnung
P	Schallleistung
p	Schalldruck
$P(\theta)$	Ausgangsleistung eines gesteuerten *Filter-and-Sum-Beamformers*
$P^{(\mathrm{DSB})}(\theta)$	Ausgangsleistung eines gesteuerten DSBs
$P^{(\mathrm{GEV})}(\theta)$	Ausgangsleistung eines GEV *Beamformer*
$P^{(\mathrm{PCA})}(\theta)$	Ausgangsleistung eines PCA *Beamformer*
$P_f(\mu_0)$	Potenzreihe der Funktion $f(\mu)$ um μ_0 herum
$P_{X_{i,m}X_{l,m}}(\Omega_k)$ ...	Kurzzeit-Kreuzleistungsdichtespektrum (Kreuzperiodogramm) des m-ten Segments zwischen $X_i(\Omega_k)$ und $X_l(\Omega_k)$
$\mathbf{p}_\kappa$	iterativer Projektionsvektor
$\mathbf{p}_n$	Position der Störquelle im Raum
$\mathbf{p}_i$	Position des i-ten Mikrophons im Raum
$\mathbf{p}_s$	Position des Sprechers im Raum
$\mathbf{p}_t$	Zielkoordinaten der Blickrichtung des *Arrays*
$p(\theta;\Omega)$	Wahrscheinlichkeitsdichtefunktion der Sprecherrichtung
$p(X(\Omega_k)\|H_0(\Omega_k))$	bedingte Verteilungsdichtefunktion gegeben eine Sprachpause
$p(X(\Omega_k)\|H_1(\Omega_k))$	bedingte Verteilungsdichtefunktion gegeben eine Sprachaktivität
r	Abstand zwischen Sender und *Array*
$r(\Omega,\mathbf{p})$	*Beamformer Response*
$r(\cdot)$	Rayleigh Quotient
r_κ	iterativ bestimmter Rayleigh Quotient
$r_{il}(n)$	verallgemeinerte Kreuzkorrelation
r_H	Hallradius
$r_{il}^{(\mathrm{GEV})}(n)$	verallgemeinerte Kreuzkorrelation für GEV-Filterkoeffizienten
$r_{il}^{(\mathrm{PCA})}(n)$	verallgemeinerte Kreuzkorrelation für PCA-Filterkoeffizienten
$\mathcal{R}(\mu^2)$	Restglieder zweiter und höherer Ordnung von μ
$s_i(n)$	(zeitdiskretes) Sprachsignal am i-ten Sensor
$S_i(\Omega)$	Sprachsignal am i-ten Sensor im Frequenzbereich
$s_c(n)$	(zeitdiskretes) Sprachsignal
$S_c(\Omega)$	Sprachsignal im Frequenzbereich
$\mathrm{si}(\cdot)$	si-Funktion $\frac{\sin(x)}{x}$
$\mathrm{SNR}_{\mathrm{Array}}(\Omega)$	frequenzabhängiges SNR am *Beamfomer*-Ausgang

$\mathrm{SNR}_{\mathrm{avg}}(\Omega)$	gemittelter geschätzter SNR-Gewinn
$\mathrm{SNR}_{\mathrm{Array}}^{(\mathrm{max})}(\Omega)$	maximal erzielbares SNR am *Beamfomer*-Ausgang
$\mathrm{SNR}_{\mathrm{Sensor},i}(\Omega)$ ..	frequenzabhängiges SNR des i-ten Sensors
$\mathrm{SNR}_{\mathrm{Sensor}}(\Omega)$ ...	frequenzabhängiges SNR gemittelt über alle Sensoren
SNRG	SNR-Gewinn
$\mathrm{SNRG}_{\kappa}(\Omega_k)$	iterativ bestimmter frequenzabhängiger asymptotischer SNR-Gewinn
$\overline{\mathrm{SNRG}}_{\kappa}$	iterativ bestimmter asymptotischer SNR-Gewinn
T	Abtastperiode
t	kontinuierliche Zeitvariable
T_{60}	Nachhallzeit
T_A	Anfangsnachhallzeit
t_g	Zeitgrenze zur Einteilung des nützlichen Schalls
T_n	Menge der Zeitindizes des Rauschsignals (ohne Sprachanteil)
T_s	Menge der Zeitindizes des Sprachsignals
$\mathbf{U}(\Omega)$	Störreferenzsignale am Ausgang der *Blocking Matrix*
$u_{s,i}(n)$	Sprachsignalkomponente am Ausgang der *Blocking Matrix*
$u_{n,i}(n)$	Störsignalkomponente am Ausgang der *Blocking Matrix*
V	Volumen eines Raums
v_g	Grenzwert für die Dominanz einer Quelle pro Frequenzkomponente
$\mathbf{v}_i$	i-ter Eigenvektor
$\hat{\mathbf{v}}_1(\Omega)$	Schätzung des dominanten Eigenvektors
$\hat{\mathbf{v}}_{1,\kappa}(\Omega)$	iterativ geschätzter dominanter Eigenvektor
$W(\Omega)$	spektrale Gewichtung
$w(\Omega)$	Nachfilter
$w_{\mathrm{BAN}}(\Omega)$	Nachfilter der blinden analytischen Normalisierung
$w_{\mathrm{GMVDR}}(\Omega)$	GMVDR-Gwichtungsfaktor
$w_{\mathrm{MN}}(\Omega)$	Nachfilter der Maximum Normalisierung
$w_{\mathrm{opt}}(\Omega)$	optimales Nachfilter
$w_{\mathrm{BSN}}(\Omega)$	Nachfilter der blinden statistischen Normalisierung
$w_{\mathrm{WPF}}(\Omega)$	Wiener *Post Filter*
$\mathbf{W}_{i,m}(\Omega_k)$	Filterkoeffizientenvektor der OPCA
$\boldsymbol{\mathcal{W}}(\Omega)$	frei wählbarer Vektor der *Blocking Matrix*
$X_i(\Omega)$	Eingangssignal am i-ten Sensor im Frequenzbereich
$x_i(n)$	(zeitdiskretes) Eingangssignal am i-ten Sensor
$Y(\Omega)$	*Beamformer*-Ausgangssignal im Frquenzbereich
$y(n)$	(zeitdiskretes) *Beamformer*-Ausgangssignal
$y_n(n)$	(zeitdiskretes) *Beamformer*-Ausgangssignal der Störkomponente
$y_s(n)$	(zeitdiskretes) *Beamformer*-Ausgangssignal der Sprachkomponente
$Y_{\mathrm{FBF}}(\Omega)$	Ausgangssignal des *Fixed Beamformers*
$Y_{\mathrm{GSC}}(\Omega)$	Ausgangssignal des *Generalized Sidelobe Cancellers*
$Y_{\mathrm{opt}}(\Omega)$	optimales Sprachsignal am *Beamformer*-Ausgang
$Y_{\mathrm{ref}}(\Omega)$	Sprachreferenzsignal am *Beamformer*-Ausgang
$\mathbf{Z}(\Omega)$	Filterkoeffizientenvektor der *Noise Cancellation*
$\mathbf{Z}_{\mathrm{opt}}(\Omega)$	optimale Filterkoeffizientenvektor der *Noise Cancellation*

Griechische Formelzeichen

α	Glättungskonstante
α_A	Absorptionsgrad einer homogenen Fläche
$\bar{\alpha}_A$	mittlerer Absorptionsgrad des Schalls für einen Raum
$\beta(\Omega)$	frequenzabhängiger Lagrange-Multiplikator
χ	Verhältnis zwischen größtem und kleinstem Eigenwert
$\delta(x)$	Delta-Distribution
$\delta_{\mathrm{LDS}}(\Omega)$	LDS-Verhältnis der reinen Sprachsignale vor und nach dem ANC
ΔBA	Unterschied der *Blocking Ability*
Δd	Abweichung des Sensorabstands
Δf	Frequenzauflösung
$\Delta\theta$	Abweichung von der Sprecherrichtung
$\Delta\Omega$	normierte Frequenzabweichung
ΔSNRG	Unterschied des SNR-Gewinns
$\Delta\sigma_N^2$	Abweichung der Rauschvarianzschätzung
$\eta(\Omega)$	skalarer komplexer frequenzabhängiger Faktor des ML-Ansatzes
η	Verhältnis von räumlich unkorreliertem zu korreliertem Rauschen
η_0	Schwellwert der VAD-Entscheidung für eine Sprachpause
η_1	Schwellwert der VAD-Entscheidung für Sprachaktivität
$\gamma(\Omega_k)$	*a posteriori* SNR
$\gamma_{X_iX_l}(\Omega)$	komplexe Kohärenzfunktion zwischen $X_i(\Omega)$ und $X_l(\Omega)$
$\Gamma_{X_iX_l}(\Omega)$	Betragsquadrat der Kohärenzfunktion zwischen $X_i(\Omega)$ und $X_l(\Omega)$
κ	Iterationsindex
λ_{min}	minimale Wellenlänge des betrachteten Wellenfeldes
$\lambda^{(\mathrm{max})}$	größter Eigenwert
$\lambda_S^{(\mathrm{max})}(\Omega)$	größter frequenzabhängiger Eigenwert (gegeben $\mathbf{\Phi_{SS}}(\Omega)$ und $\mathbf{\Phi_{NN}}(\Omega)$)
$\lambda_X^{(\mathrm{max})}(\Omega)$	größter frequenzabhängiger Eigenwert (gegeben $\mathbf{\Phi_{XX}}(\Omega)$ und $\mathbf{\Phi_{NN}}(\Omega)$)
λ_i	i-ter Eigenwert
$\mathbf{\Lambda}$	Diagonalmatrix der Eigenwerte
μ	Schrittweite
$\mu^{(\mathrm{Neu})}$	Schrittweite der neuen Adaptionsregel
$\mu^{(\mathrm{Oja})}$	Schrittweite der Oja-Adaptionsregel
ν	Laufvariable
Ω	normierte kontinuierliche Kreisfrequenz
Ω_k	k-te normierte diskrete Kreisfrequenz
P	Anzahl der Nutzsignalquellen bei der BSS
$\hat{\phi}^{(\mathrm{GG})}_{Y_{\mathrm{FBF}}Y_{\mathrm{FBF}}}(\Omega)$ ...	spektrale Leistungsdichte am Ausgang des FBFs
$\hat{\phi}^{(\mathrm{GG})}_{Y_{\mathrm{GSC}}Y_{\mathrm{GSC}}}(\Omega)$ ...	spektrale Leistungsdichte am Ausgang des GSCs
$\phi_{X_iX_l}(\Omega)$	Kreuzleistungsdichtespektrum zwischen $X_i(\Omega)$ und $X_l(\Omega)$
$\phi_{YY}(\Omega)$	LDS des *Beamformer*-Ausgangssignals
π	3,14159265359...
$\mathbf{\Phi_{NN}}(\Omega)$	Matrix der spektralen Kreuzleistungsdichten der Störsignale $\mathbf{N}$
$\mathbf{\Phi_{XX}}(\Omega)$	Matrix der spektralen Kreuzleistungsdichten der Mikrophonsignale $\mathbf{X}$
$\mathbf{\Phi_{SS}}(\Omega)$	Matrix der spektralen Kreuzleistungsdichten der Sprachsignale $\mathbf{S}$
$\mathbf{\Phi}^{(\mathbf{XN})}$	Kombination der KLDS-Matrizen von Stör- und Sprachsignalen

$\widetilde{\mathbf{\Phi}}_{\mathbf{NN}}$	Normierte Matrix der spektralen Kreuzleistungsdichten der Störsignale
$\varpi_{i,m}^{(\mathrm{BM})}(\Omega_k)$	i-te binäre Maske der BSS für den Block m
$\varpi_{i,m}^{(\mathrm{LM})}(\Omega_k)$	i-te *Likelihood Maske* der BSS für den Block m
$\Psi_m(\Omega_k)$	geglättete Entscheidungsvariable für den Block m
$\hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}^{(\mathrm{GG})}$	iterative Schätzung der KLDS-Matrix durch eine gleichmäßige Gewichtung
$\hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}^{(\mathrm{EG})}$	iterative Schätzung der KLDS-Matrix durch eine exponentielle Glättung
$\hat{\mathbf{\Phi}}_{\mathbf{XX},\kappa}^{(\mathrm{IS})}$	iterative Schätzung der KLDS-Matrix durch eine instantane Schätzung
ρ_R	Schall-Reflexionsgrad einer homogenen Fläche
ρ	Schrittweitefaktor
σ^2	Varianz
σ^2_{LDS}	Varianz der LDS-Verhältnise
$\sigma^2_N(\Omega_k)$	Varianz des Störsignals $N(\Omega_k)$
$\sigma^2_S(\Omega_k)$	Varianz des Sprachsignals $S(\Omega_k)$
τ	zeitliche Dämpfungskonstante
τ_e	effektive Zeitverzögerung
τ_g	zeitliche Einwirktiefe einer Glättung
τ_i	Laufzeit des Signals von der Quelle bis zum i-ten Mikrophon
τ_{il}	die Zeitverzögerung zwischen zwei Signalen $x_i(t)$ und $x_l(t)$
θ	Winkel
θ_n	Richtung der Störschallquelle
$\theta_{n,i}$	Richtung der i-ten Störschallquelle
θ_s	Sprecherrichtung
θ_{s1}	Sprecherrichtung des ersten Sprechers für die BSS
θ_{s2}	Sprecherrichtung des zweiten Sprechers für die BSS
θ_t	Richtungswinkel des *Arrays* bezüglich eines Ziels
$\vartheta(\cdot)$	relativer Anteil an nützlichem Schall
$\theta_{t\,\nu}$	diskretisierte Zielrichtungen
ξ	Zielfunktion
$\xi(\Omega_k)$	*a priori* SNR
$\tilde{\xi}$	fehlerhafte Schätzung des *a priori* SNRs
$\zeta(\Omega)$	komplexer Skalar

Spezielle Symbole

$*$	Faltungsoperator
$(\cdot)^*$	konjugiert komplexe Schreibweise
$(\cdot)^H$	hermitesch konjungierte Notation
$(\cdot)^T$	transponierte Schreibweise
$\mathbf{I}_M$	Einheitsmatrix der Dimension M
$\Im\{\cdot\}$	Imaginärteil
$\nabla_{\mathbf{F}}\{\cdot\}$	Ableitung bezüglich eines komplexen Vektors
$\frac{\partial}{\partial \mathbf{F}}$	komplex konjugierte Ableitung des Vektors $\mathbf{F}$
$\Re\{\cdot\}$	Realteil

Rang($\mathbf{A}$)	Rang der Matrix $\mathbf{A}$
Spur($\mathbf{A}$)	Spur der Matrix $\mathbf{A}$
MAX$\{\cdot\}$	Maximum-Operator
var$\{\cdot\}$	Varianz
$\|\| \cdot \|\|$	L_2-Norm

Abkürzungen

AMCor	*Amplituden Modulation Correlation*
ANC	*Adaptive Noise Cancellation*
ASC	*Adaptive Speech Cancellation*
ASCBM	*Adaptive Speech Cancellation Blocking Matrix*
BA	*Blocking Ability*
BAN	blinde analytische Normalisierung
BM	*Binary Masking*
BM	*Blocking Matrix*
BSS	*Blind Source Separation*
DD	*Decision-Directed*
DFT	*Discrete Fourier Transform*
DI	*Directivity Index*
DO	*Delay Only*
DOA	*Direction-of-Arrival*
DOBM	*Delay Only Blocking Matrix*
DOR	*Delay Only Ratio*
DORBM	*Delay Only Ratio Blocking Matrix*
DR	*Distortionless Response*
DSB	*Delay-and-Sum-Beamformer*
DTFT	*Discrete Time Fourier Transform*
DUET	*Degenerate Unmixing Estimation Technique*
EDC	*Energy Decay Curve*
EDT	*Early Decay Time*
EG	Exponentielle Gewichtung
FBF	*Fixed Beamformer*
FEM	*Finite Element Method*
FFT	*Fast Fourier Transform*
FIR	*Finite Impulse Response*
FSB	*Filter-and-Sum-Beamformer*
GCC	*Generalized Cross Correlation*
GEV	*Generalized Eigenvector*
GEVBM	*Generalized Eigenvector Blocking Matrix*
GEVP	*Generalized Eigenvalue Problem*
GG	Gleichmäßige Gewichtung
GML	*Generalized Maximum Likelihood*
GMMSE	*Generalized Minimum Mean Squared Error*
GMV	*Generalized Minimum Variance*
GMVDR	*Generalized Minimum Variance Distortionless Response*

GSC	*Generalized Sidelobe Canceller*
GSVD	*Generalized Singular Value Decomposition*
HMM	*Hidden Markov Modell*
ICA	*Independent Component Analysis*
ICMA	*In Situ Calibrated Microphone Array*
IDFT	*Inverse Discrete Fourier Transform*
IFFT	*Inverse Fast Fourier Transform*
IS	Instantaner Schätzer
ITU	*International Telecommunication Union*
KLDS	Kreuzleistungsdichespektrum
LCMVDR	*Linearly Constrained Minimum Variance Distortionless Response*
LDS	Leistungsdichtespektrum
LE	*Localization Error*
LM	*Likelihood Masking*
LMS	*Least Mean Squares*
LOS	*Line of Sight*
LRT	*Likelihood Ratio Test*
LSE	*Least Squares Error*
MAP	maximum *a posteriori*
MCWF	*Multi Channel Wiener Filter*
MF	*Matched Filter*
MFB	*Matched Filter Beamformer*
MIMO	*Multiple Input Multiple Output*
ML	*Maximum Likelihood*
ML-STBF.....	*Maximum Likelihood Steered Adaptive Beamformer*
MMSE	*Minimum Mean Squared Error*
MN	Maximum-Normalisierung
MOP	*Mutual Orthogonal Projection*
MOS	*Mean Opinion Score*
WPF	*Wiener Post Filter*
MS	Minimum Statistik
MSC	*Magnitude Squared Coherence*
MUSIC	*Multiple Signal Classification*
MV	*Minimum Variance*
MWF	*Multi Channel Wiener Filter*
NC	*Noise Cancellation*
OCPCA	*Orthogonal Constraint Principal Component Analysis*
ODG	*Objective Difference Grade*
PAST	*Projection Approximation Subspace Tracking*
PC	*Personal Computer*
PCA	*Principal Component Analysis*
PDF	*Probability Density Function*
PESQ	*Perceptual Evaluation of Speech Quality*
PHAT	*Phase Transform*
PSM	*Perceptual Similarity Measure*
RIA	Raumimpulsantwort

RLS	*Recursive Least Squares*
ROC	*Receiver Operating Characteristic*
SAD	*Symmetric Adaptive Decorrelation*
SIR	*Signal-to-Interference-Ratio*
SNR	*Signal-to-Noise-Ratio*
SRP	*Steered Response Power*
TDOA	*Time-Difference of Arrival*
TF	*Transfer Function*
TFBM	*Transfer Function Blocking Matrix*
TFR	*Transfer Function Ratio*
TFRBM	*Transfer Function Ratio Blocking Matrix*
VAD	*Voice Activity Detection*

Literaturverzeichnis

[AB79] Allen, J. B. ; Berkley, D. A.: Image Method for Efficiently Simulating Small-Room Acoustics. In: *Journal of the Acoustical Society of America* 107 (1979), Nr. 4, S. 943–950

[AG96] Affes, S. ; Grenier, Y.: A Source Subspace Tracking Array of Microphones for Double Talk Situations. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)* Bd. 2. Atlanta, USA, May 1996, S. 909–912

[AG97] Affes, S. ; Grenier, Y: A Signal Subspace Tracking Algorithm for Microphone Array Processing of Speech. In: *IEEE Transactions on Speech and Audio Processing* 5 (1997), Sept., S. 425–437

[AHBK03] Aichner, R. ; Herbordt, W. ; Buchner, H. ; Kellermann, W.: Least-Squares Error Beamforming using Minimum Statistics and Multichannel Frequencydomain Adaptive Filtering. In: *Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*. Kyoto, Japan, Sept. 2003, S. 223–226

[AK00] Anemüller, J. ; Kollmeier, B.: Amplitude Modulation Decorrelation for Convolutive Blind Source Separation. In: *Proc. of the second international workshop on independent component analysis and blind signal separation*. Helsinki, Finland, June 2000, S. 215–220

[AL05] Alli, M. ; Lyons, R.: A Root of less Evil. In: *IEEE Signal Processing Magazine* 9 (2005), S. 58–67

[Ama77] Amari, S.: Neural theory of association and concept-formation. In: *Biological Cybernetics* 26 (1977), Sept., Nr. 3, S. 175–185

[Ami] *Amigo - Ambient Intelligence for the Networked Home Environment*. `http://www.amigo-project.org`

[Ane01] Anemüller, J.: *Across-Frequency Processing in Convolutive Blind Source Separation*, University of Oldenburg, Germany, Diss., 2001

[Arn51] Arnoldi, W. E.: The Principle of Minimized Iterations in the Solution of the Matrix Eigenvalue Problem. In: *Quarterly of Applied Mathematics* (1951), 9, S. 17–29

[Bar03] Bartsch, G.: *Effiziente Methoden für die niederfrequente Schallfeldsimulation*, RWTH Aachen, Germany, Diss., 2003

[BCM05] Benesty, J. ; Chen, J. ; Makino, S.: *Speech Enhancement*. Springer-Verlag, 2005

[Ber96] Beranek, L.: Concert and Opera Halls: How They Sound. In: *Acoustical Physics* 42 (1996), S. 779–780

[Bit02] Bitzer, J.: *Mehrkanalige Geräuschunterdrückungssysteme - eine vergleichende Analyse*, Universität Bremen, Germany, Diss., 2002

[Bod56] Bodewig, E.: *Matrix Calculus*. North-Holland, Amsterdam, 1956

[BP66] Bendat, J. S. ; Piersol, A. G.: *Measurement and Analysis of Random Data*. New York : Wiley, 1966

[BP80] Bendat, J. S. ; Piersol, A. G.: *Engineering Application of Correlation and Spectral Analysis*. New York : Wiley, 1980

[Bra99] Brandstein, M.: Time-Delay Estimation of Reverberated Speech Exploiting Harmonic Structure. In: *Journal of the Acoustical Society of America* 105 (1999), May, S. 2914–2919

[BS73] Bangs, W. J. ; Schultheiss, P. M.: Space Time Processing for Optimal Parameter Estimation. In: *Signal Processing* (1973), S. 577–590

[BS01] Bitzer, J. ; Simmer, K. U.: Superdirective Microphone Arrays. In: Brandstein, M.S. (Hrsg.) ; Ward, D.B. (Hrsg.): *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag, 2001, S. 19–38

[BSK99a] Bitzer, J. ; Simmer, K. ; Kammeyer, K.: An Alternative Implementation of the Superdirective Beamformer. In: *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. New Paltz NY, USA, 1999, S. 7–10

[BSK99b] Bitzer, J. ; Simmer, K. U. ; Kammeyer, K.-D.: Multi-Microphone Noise Reduction by Post-Filter and Superdirective Beamformer. In: *Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*. Pocono Manor, USA, Sept. 1999, S. 100–103

[BSK99c] Bitzer, J. ; Simmer, K. U. ; Kammeyer, K.-D.: Theoretical Noise Reduction Limits of the Generalized Sidelobe Canceller (GSC) for Speech Enhancement. In: *ICASSP* Bd. 4. Phoenix, Arizona, March 1999, S. 2965–2968

[BSRG05] Bhiksha, R. ; Seltzer, M. ; Reyes-Gomez, M. J.: Speech Recognizer based Maximum Likelihood Beamforming. In: Divenyi, P. (Hrsg.): *Speech Separation by Humans and Machines*. Springer US, 2005, S. 65–82

[Buc07] Buck, M.: Optimaler Beamformer-Entwurf unter Berücksichtigung spezifischer Mikrofoneigenschaften. In: *Fortschritte der Akustik - DAGA 2007, DEGA e.V.* Stuttgart, März 2007, S. 335–336

[CA03] Cichocki, A. ; Amari, S.: *Adaptive Blind Signal and Image Processing*. John Wiley & Sons, 2003

[Cap69] Capon, J.: High-Resolution Frequency-Wavenumber Wpectrum Analysis. In: *Proceedings of the IEEE* (1969), Aug., S. 1408–1418

[CBHD06] Chen, J. ; Benesty, J. ; Huang, Y. ; Doclo, S.: New Insights into the Noise Reduction Wiener Filter. In: *IEEE Transactions on Audio, Speech and Language Processing* 14 (2006), July, S. 1218– 1234

[CHY98] Chen, T. ; Hua, Y. ; Yan, W. Y.: Global Convergence of Oja's Subspace Algorithm for Principal Component Extraction. In: *Journal of Mathematical Analysis and Applications* 106 (1998), S. 69–84

[CK01] Cho, Y.D. ; Kondoz, A.: Analysis and Improvement of a Statistical Model-based Voice Activity Detector. In: *IEEE Signal Processing Letters* 8 (2001), Oct., S. 276–278

[CM78] Cremer, L. ; Müller, H. A.: *Die wissenschaftlichen Grundlagen der Raumakustik. Band I.* S. Hirzel, 1978

[CWB+55] Cook, R. K. ; Waterhouse, R. V. ; Berendt, R. D. ; Edelman, S. ; Thompson, M. C.: Measurement of Correlation Coefficients in Reverberant Sound Fields. In: *Journal Acoust. Soc. Am.* 27 (1955), Nr. 6, S. 1072–1077

[CZK86] Cox, H. ; Zeskind, R. ; Kooij, T.: Practical Supergain. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 34 (1986), June, Nr. 3, S. 393– 398

[CZO87] Cox, H. ; Zeskind, R. M. ; Owen, M. M.: Robust Adaptive Beamforming. In: *IEEE Transactions on Acoustics, Speech, Signal Processing* 35 (1987), Oct., S. 1365–1376

[DCP01] Di Claudio, E. D. ; Parisi, R.: Multi-Source Localization Strategies. In: Brandstein, M.S. (Hrsg.) ; Ward, D.B. (Hrsg.): *Microphone Arrays: Signal Processing Techniques and Applications.* Springer-Verlag, 2001, S. 181–201

[DCP03] Di Claudio, E. D. ; Parisi, R.: Robust ML Wideband Beamforming in Reverberant Fields. In: *IEEE Transactions on Signal Processing* 51 (2003), Feb., S. 338 –349

[DDP88] Dal-Degan, N. ; Prati, C.: Acoustic Noise Analysis and Speech Enhancement Techniques for Mobile Radio Applications. In: *Signal Processing* 15 (1988), Nr. 4, S. 43–56

[DFG01] Doucet, A. ; Freitas, N. de ; Gordon, N.: *Sequential Monte Carlo Methods in Practice.* Springer-Verlag, 2001

[Dic97] Dickreiter: *Handbuch der Tonstudiotechnik.* München : Sauerverlag KG, 1997

[DK96] Diamantaras, K. I. ; Kung, S. Y.: *Principal Component Neural Networks - Theory and Applications.* John Wiley & Sons, 1996

[DM99] Doclo, S. ; Moonen, M.: Robustness of SVD-based Optimal Filtering for Noise Reduction in Multi-Microphone Speech Signals. In: *Proc. of the 1999 IEEE International Workshop on Acoustic Echo and Noise Control (IWAENC'99).* Pocono Manor, Pennsylvania, USA, Sep. 1999, S. 80–83

[DM01] Doclo, S. ; Moonen, M.: GSVD-based Optimal Filtering for Multi-Microphone Speech Enhancement. In: Brandstein, M.S. (Hrsg.) ; Ward, D.B. (Hrsg.): *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag, 2001, S. 111–132

[DM05] Doclo, S. ; Moonen, M.: Multimicrophone Noise Reduction using Recursive GSVD-based Optimal Filtering with ANC Postprocessing Stage. In: *IEEE Transactions on Speech and Audio Processing* 13 (2005), Jan., S. 53– 69

[DM06] Doclo, S. ; Moonen, M.: Superdirective Beamforming Robust Against Microphone Mismatch. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. Toulouse, France, May 2006, S. 41–44

[Dob06] Doblinger, G.: An adaptive Microphone Array for optimum Beamforming and Noise Reduction. In: *Proc. European Signal Processing Conference (EUSIPCO)* Bd. 2. Florence, Italy, May 2006

[DPK96] Dau, T. ; Puschel, D. ; Kohlrausch, A.: A Quantitative Model of the Effective Signal Processing in the Auditory System. In: *Journal of the Acoustical Society of America* 99 (1996), Nr. 6, S. 3615–3622

[Dre99] Drews, M.: *Mikrofonarrays und mehrkanalige Signalverarbeitung zur Verbesserung gestörter Sprache*, Technische Universität Berlin, Germany, Diss., 1999

[DSB01] DiBiase, J. ; Silverman, H. ; Brandstein, M.: Robust Localization in Reverberant Rooms. In: Brandstein, M.S. (Hrsg.) ; Ward, D.B. (Hrsg.): *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag, 2001, S. 157–180

[DSWM05] Doclo, S. ; Spriet, A. ; Wouters, J. ; Moonen, M.: Speech Distortion Weighted Multichannel Wiener Filtering Techniques for Noise Reduction. In: Benesty, J. (Hrsg.) ; Huang, A. (Hrsg.) ; S., Makino (Hrsg.): *Speech Enhancement*. Springer-Verlag, 2005, S. 199–228

[EK03] Elmusrati, M. ; Koivo, H.: Multi-Path MVDR Smart Antenna Algorithm for Frequency Selective Channels. In: *Proc. Int. ITG-Conf. on Antennas (INICA)*. Berlin, 2003, S. 369–371

[Elk00] Elko, G. W.: Superdirectional Microphone Arrays. In: Gay, S. L. (Hrsg.) ; Benesty, J. (Hrsg.): *Acoustic Signal Processing for Telecommunication*. Kluwer Academic Publishers, 2000, S. 181–237

[EM84] Ephraim, Y. ; Malah, D.: Speech Enhancement using a Minimum Mean Square Error Short-Time Spectral Amplitude Estimator. In: *IEEE Transactions on Acoust., Speech, Signal Processing* ASSP-32 (1984), Dec., S. 1109–1121

[ETS02] ETSI: *Speech Processing, Transmission and Quality Aspects; Distributed Speech Recognition; advanced front-end feature extraction algorithm; compression algorithms*. 2002. – ETSI ES 201 108 Recommendation

[Eyr30] Eyring, C. F.: Reverberation time in "dead" rooms. In: *Journal of the Acoustical Society of America* (1930), S. 217–241

[Fis07] Fischer, C.: *Realisierung eines akustischen Beamformings unter Verwendung von Verfahren zur adaptiven Eigenwertzerlegung.* 2007. – Studienarbeit, Fachgebiet Nachrichtentechnik, Universität Paderborn

[Flo01] Florencio, H. S.: Multichannel Filtering for optimum Noise Reduction in Microphone Arrays. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP).* Salt Lake City, USA, May 2001, S. 197–200

[Fra61] Francis, J. G. F.: The QR Transformation: A Unitary Analogue to the LR Transformation, PartI and Part II. In: *The Computer Journal* (1961), 4, S. 265–272, 332–345

[Fro72] Frost, O. L.: An Algorithm for Linearly Constrained Adaptive Array Processing. In: *Proceedings of the IEEE* 60 (1972), August, Nr. 8, S. 926–935

[FSJ93] Flanagan, J. L. ; Surendran, A. C. ; Jan, E. E.: Spatially Selective Sound Capture for Speech and Audio Processing. In: *Speech Communication* 13 (1993), Oct., S. 207–222

[GAG96] Gazor, S. ; Affes, S. ; Grenier, Y.: Robust Adaptive Beamforming via Target Tracking. In: *IEEE Transactions on Signal Processing* 44 (1996), June, S. 1589–1593

[Gan00] Gannot, S.: *Array Processing of Nonstationary Signals with Application to Speech*, Tel-Aviv University, Israel, Diss., 2000

[Gar92] Gardner, W. A.: A Unifying View of Coherence in Signal Processing. In: *Signal Processing* 29 (1992), Nr. 2, S. 113–140

[GBW99] Gannot, S. ; Burshtein, D. ; Weinstein, E.: Beamforming Methods for Multi-Channel Speech Enhancement. In: *Int. Workshop on Acoustic Echo and Noise Control (IWAENC).* Pocono Manor, USA, Sept. 1999, S. 96–99

[GBW01] Gannot, S. ; Burshtein, D. ; Weinstein, E.: Signal Enhancement using Beamforming and Nonstationarity with Applications to Speech. In: *IEEE Transactions on Signal Processing* 49 (2001), Aug., Nr. 8, S. 1614–1626

[GBW04] Gannot, S. ; Burshtein, D. ; Weinstein, E.: Analysis of the Power Spectral Deviation of the General Transfer Function GSC. In: *IEEE Transactions on Signal Processing* 52 (2004), April, S. 1115–1121

[GC04] Gannot, S. ; Cohen, I: Speech Enhancement based on the General Transfer Function GSC and Postfiltering. In: *IEEE Transactions on Speech and Audio Processing* 12 (2004), Nov., Nr. 6, S. 561–571

[GJ82] Griffiths, L. J. ; Jim, C. W.: An Alternative Approach to Linearly Constrained Adaptive Beamforming. In: *IEEE Trans. on Antennas and Propagation* 30 (1982), January, Nr. 1, S. 27–34

[GM55] Gilbert, E.N. ; Morgan, S.P.: Optimum Design of Directive Antenna Arrays Subject to Random Variables. In: *Bell Systems Technical Journal* 34 (1955), May, S. 637–663

[GM76] Gray, A. ; Markel, J.: Distance Measures for Speech Processing. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 24 (1976), Oct., Nr. 8, S. 380–391

[GN02] Grbić, N. ; Nordholm, S.: Soft Contrained Subband Beamforming for Hands-Free Speech Enhancement. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. Orlando, USA, May 2002, S. 885–888

[Gri67] Griffiths, L. J.: A comparison of multidimensional Wiener and maximum-likelihood filters for antenna arrays. In: *IEEE Proceedings* 55 (1967), Nov., S. 2045– 2047

[GRT03] Gustafsson, T. ; Rao, B. D. ; Trivedi, M.: Source Localization in Reverberant Environments: Modeling and Statistical Analysis. In: *IEEE Transactions on Speech and Audio Processing* 11 (2003), Nov., S. 791–803

[GV99] Golub, G. ; Vorst, H. van d.: *Numerical Progress in Eigenvalue Computation in the 20th Century*. `citeseer.ist.psu.edu/golub99numerical.html`. Version: 1999

[GV00] Golub, G. H. ; Vorst, H. A. d.: Eigenvalue Computation in the 20th Century. In: *Journal of Computational and Applied Mathematics* 123 (2000), Nov., Nr. 1-2, S. 35–65

[Has02] Hasu, V.: Eigenvalue Approach to Joint Power Control and Beamforming for CDMA Systems. In: *IEEE Seventh International Symposium on Spread Spectrum Techniques and Applications (ISSSTA)*. Prague, Czech, Sept. 2002, S. 561–565

[Hay02] Haykin, S.: *Adaptive Filter Theory*. Prentice Hall, 2002

[HBD00] Hammerschmidt, J. S. ; Brunner, C. ; Drewes, C.: Eigenbeamforming – A Novel Concept in Array Signal Processing. In: *Proc. of European Wireless Conference*. Dresden, Germany, Sept. 2000

[HBNK07] Herbordt, W. ; Buchner, H. ; Nakamura, S. ; Kellermann, W.: Multichannel bin-wise robust frequency-domain adaptive filtering and its application to adaptive beamforming. In: *IEEE Transactions on Audio, Speech and Language Processing* 15 (2007), May, Nr. 4, S. 1340–1351

[Her04] Herbordt, W.: *Combination of Robust Adaptive Beamforming with Acoustic Echo Cancellation for Acoustic Human/Machine Interfaces*, Universität Erlangen-Nuremberg, Germany, Diss., 2004

[HGJ06] Hongqing, l. ; Guisheng, L. ; Jie, Z.: A robust adaptive Capon beamforming. In: *Signal Processing* 86 (2006), Oct., S. 2820–2826

[HK00] Hansen, M. ; Kollmeier, B.: Objective Modeling of Speech Quality with a Psychoacoustically Validated Auditory Model. In: *Journal Audio Eng. Soc.* 48 (2000), Nr. 5, S. 395–409

[HK01] Herbordt, W. ; Kellermann, W.: Efficient Frequency-Domain Realization of Robust Generalized Sidelobe Cancellers. In: *IEEE Workshop on Multimedia Signal Processing (MMSP)*. Cannes, Oct. 2001

[HK02] Herbordt, W. ; Kellermann, W.: Analysis of Blocking Matrices for Generalized Sidelobe Cancellers for Non-Stationary Broadband Signals. In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*. Orlando, USA, May 2002

[HK03] Herbordt, W. ; Kellermann, W.: Adaptive Beamforming for Audio Signal Acquisition. In: Benesty, J. (Hrsg.) ; Huang (Hrsg.): *Adaptive Signal Processing*. Springer-Verlag, 2003, S. 155–194

[HKO01] Hyvärinen, A. ; Karhunen, J. ; Oja, E.: *Independent Component Analysis*. John Wiley & Sons, 2001

[HN76] Hodgkiss, W. S. ; Nolte, L. W.: Covariance between Fourier Coefficients representing the Time Waveforms observed from an Array of Sensors. In: *Journal of the Acoustical Society of America* 59 (1976), March, S. 582–590

[Hou64] Housholder, A. S.: *The Theory of Matrices in Numerical Analysis*. Dover, New York, 1964

[HS01] Hoshuyama, O. ; Sugiyama, A.: Robust adaptive beamforming. In: Brandstein, M.S. (Hrsg.) ; Ward, D.B. (Hrsg.): *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag, 2001, S. 19–38

[HSH96] Hoshuyama, O. ; Sugiyama, A. ; Hirano, A.: A Robust Adaptive Beamformer for Microphone Arrays with a Blocking m Matrix using Constrained Adaptive Filters. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. Atlanta, USA, May 1996, S. 925–928

[HSH99] Hoshuyama, O. ; Sugiyama, A. ; Hirano, A.: A Robust Adaptive Beamformer for Microphone Arrays with a Blocking Matrix using Constrained Adaptive Filters. In: *IEEE Transactions on Signal Processing* 47 (1999), Oct., S. 2677–2684

[HT73] Hahn, W. ; Tretter, S.: Space Time Processing for Optimal Parameter Estimation. In: *IEEE Transactions on Information Theory* 19 (1973), Sept., S. 608–614

[Hub03] Huber, R.: *Objective Assessment of Audio Quality using an Auditory Processing Model*, University of Oldenburg, Germany, Diss., 2003

[Hub06] Huber, R.: Vorhersage der empfundenen Klangqualität von Mehrkanal-Störgeräuschreduktionsverfahren in Personenkraftwagen. In: *Fortschritte der Akustik - DAGA 2006, DEGA e.V.* Berlin, März 2006, S. 219–220

[HUKW08] Häb-Umbach, R. ; Krüger, A. ; Warsitz, E.: Blinde akustische Strahlformung für Anwendungen im KFZ. In: *Fortschritte der Akustik - DAGA 2008, DEGA e.V.* Dresden, März 2008

[HUW05] Haeb-Umbach, R. ; Warsitz, E.: Adaptive Filter-and-Sum Beamforming in Spatially Correlated Noise. In: *Int. Workshop on Acoustic Echo and Noise Control (IWAENC).* Eindhoven, Netherlands, Sept. 2005

[IEE79] *Programs for Digital Signal Processing.* John Wiley & Sons, 1979. – IEEE Press, Chap. 8.1

[IN06] Ishizuka, K. ; Nakatani, T.: Study of Noise Robust Voice Activity Detection based on Periodic Component to Aperiodic Component Ratio. In: *Statistical And Perceptual Audition (SAPA).* Pittsburgh, USA, Sept. 2006

[Iri97] Irie, R. E.: Multimodal Sensory Integration for Localization in a Humanoid Robot. In: *Proc. of Second IJCAI Workshop on Computational Auditory Scene Analysis (CASA97).* Nagoya, Japan, Aug. 1997, S. 54–58

[IS70] Itakura, F. ; Saito, S.: A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies. In: *Electronics and Communications in Japan* 53A (1970), S. 36–43

[ITU01] ITU: Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs. In: *Series P: Telephone Transmission Quality Recommendation P.862.* International Telecommunications Union (ITU), 2001

[Jac46] Jacobi, C. G. J.: Über ein leichtes Verfahren die in der Theorie der Säculärstörungen vorkommenden Gleichungen numerisch aufzulösen. In: *Journal für die reine und angewandte Mathematik* 30 (1846), Nov., Nr. 1-2, S. 51–94

[JD93] Johnson, D. H. ; Dudgeon, D. E.: *Array Signal Processing.* New Jersey : Prentice Hal, 1993

[JF96] Jan, E. E. ; Flanagan, J.: Sound Capture from Spatial Volumes: Matched-Filter Processing of Microphone Arrays having Randomly Distributed Sensors. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP).* Atlanta, USA, 1996

[JG00] James, G. R. ; G., Rafik A.: Optimum Near-Field Performance of Microphone Arrays subject to a Far-Field Beampattern Constraint. In: *The Journal of the Acoustical Society of America* 108 (2000), Nov., S. 2248–2255

[JHLCCC06] Ju-Hong Lee, J.-H. ; Cheng, K.-P. ; C.-C., Wang: Robust Adaptive Array Beamforming under Steering Angle Mismatch. In: *Signal Processing* 86 (2006), Feb., S. 296 – 309

[JN87] Jacobsen, F. ; Nielsen, T. G.: Spatial Correlation and Coherence in a Reverberant Sound Field. In: *Journal of Sound Vibration* 118 (1987), Oct., S. 175–180

[Jor74] Jordan, W.: *47. Conventions AES.* Copenhagen : Audio Engineering Society (AES), 1974

[JRY00] Jourjine, A. ; Rickard, S. ; Yilmaz, O.: Blind Separation of Disjoint Orthogonal Signals. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. Istanbul, Turkey, June 2000, S. 2985–2988

[KAM07] Knaak, M. ; Araki, S. ; Makino, S.: Geometrically constrained Independent Component Analysis. In: *IEEE Transactions on Audio, Speech and Language Processing* 15 (2007), Feb., S. 715–726

[Kar84] Karhunen, J.: Adaptive Algorithms for Estimating Eigenvectors of Correlation Type Matrices. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)* Bd. 9. San Diego, USA, March 1984, S. 592 – 595

[KC76] Knapp, C. H. ; Carter, G. C.: The generalized correlation method for estimation of time delay. In: *IEEE Trans. ASSP* (1976), S. 320–327

[KDO05] Kristjansson, T. ; Deligne, S. ; Olsen, P.: Joint Speaker Segmentation, Localization and Identification for Streaming Audio. In: *Proc. Interspeech*. Lisbon, Portugal, Sept. 2005

[KHJ06] Kim, L.H. ; Hasegawa-Johnson, M.: Generalized optimal Multi-Microphone Speech Enhancement using sequential Minimum Variance Distortionless Response (MVDR) Beamforming and Postfiltering. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. Toulouse, France, May 2006, S. 65–68

[KK02] Kammeyer, K.D. ; Kroschel, K.: *Digitale Signalverarbeitung*. 5. Auflage. Stuttgart : Teubner, 2002

[Krü07] Krüger, A.: *Mehrkanalige Sprachsignalverbesserung mittels adaptiver Eigenwertzerlegung in einer Generalized Sidelobe Canceller Anordnung*. 2007. – Diplomarbeit 5/06, Fachgebiet Nachrichtentechnik, Universität Paderborn

[Kut00] Kuttruff, H.: *Room Acoustics*. 4th edition. Taylor & Francis Group, 2000

[Lan50] Lanczos, C.: An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators. In: *Journal of Research of the National Bureau of Standards* 45 (1950), Oct., Nr. 4, S. 255–282

[Lev64] Levin, M. J.: *Maximum-Likelihood Array Processing*. M.I.T. Lincoln Laboratory, Lexington, USA, Dec. 1964. – Technical Report DDC 455743

[LNO00] Lourens, T. ; Nakadai, K. ; Okuno, H.: Humanoid Active Audition System. In: *Proc. of First IEEE-RAS International Conference on Humanoid Robots (Humanoids2000)*. Cambridge, USA, Sep. 2000

[Loi07] Loizou, P.: *Speech Enhancement: Theory and Practice*. CRC Press, 2007

[LS05] Li, J. ; Stoica, P.: *Robust Adaptive Beamforming*. Wiley, 2005

[LV06] Lotter, T. ; Vary, P.: Dual-Channel Speech Enhancement by Superdirective Beamforming. In: *EURASIP Journal on Applied Signal Processing* 2006 (2006), S. Article ID 63297, 14 pages. – doi:10.1155/ASP/2006/63297

[LVKL96] Laakso, T. I. ; Välimäki, V. ; Karjalainen, M. ; Laine, U. K.: Splitting the Unit Delay. In: *IEEE Signal Processing Magazine* 13 (1996), Jan., Nr. 1, S. 30–60

[LWW03] Lehmann, E. A. ; Ward, D. B. ; Williamson, R. C.: Experimental Comparison of Particle Filtering Algorithms for Acoustic Source Localization in Reverberant Room. In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*. Hong Kong, China, April 2003

[MA04] Mungamuru, B. ; Aarabi, P.: Source Localization in Reverberant Environments: Modeling and Statistical Analysis. In: *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics* 34 (2004), June, S. 1526–1540

[Mak03] Makino, S.: Blind Source Separation of Convolutive Mixtures of Speech. In: Benesty, J. (Hrsg.) ; Huang (Hrsg.): *Adaptive Signal Processing*. Springer-Verlag, 2003, S. 195–225

[Mar94] Martin, R.: Spectral Subtraction based on Minimum Statistics. In: *European Signal Processing Conference (EUSIPCO)*. Edinburgh, Scotland, Sept. 1994, S. 1182–1185

[Mar95] Martin, R.: *Freisprecheinrichtungen mit mehrkanaliger Echokompensation und Störgeräuschreduktion*, Technische Hochschule Aachen, Germany, Diss., 1995

[Mar01] Martin, R.: Noise Power Spectral Density Estimation based on Optimal Smoothing and Minimum Statistics. In: *IEEE Transactions Speech and Audio Processing* 108 (2001), July, S. 504–512

[MB02] McCowan, I.A. ; Bourlard, H.: Microphone Array Post-Filter for Diffuse Noise Field. In: *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. Orlando, USA, May 2002, S. 905–908

[MB03] McCowan, I.A. ; Bourlard, H: Microphone Array Post-Filter based on Noise Field Coherence. In: *IEEE Transactions on Speech and Audio Processing* 11 (2003), S. 240–259

[MK02] Marzinzik, M. ; Kollmeier, B.: Speech Pause Detection for Noise Spectrum Estimation by Tracking Power Envelope Eynamics. In: *IEEE Transactions on Speech and Audio Processing* 10 (2002), Feb., S. 109–118

[MMM00] McCowan, I. ; Marro, C. ; Mauuary, L.: Robust Speech Recognition Using Near-Field Superdirective Beamforming with Post-Filtering. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. Istanbul, Turkey, June 2000, S. 1723 1726

[MMU98] Marro, C. ; Mahieux, Y. ; U., Simmer K.: Analysis of Noise Reduction and Dereverberation Techniques based on Microphone Arrays with Postfiltering. In: *IEEE Transactions Speech, Audio Processing* 6 (1998), May, S. 240–259

[Mor04] Morgan, D.: Adaptive Algorithms for solving Generalized Eigenvalue Signal Enhancement Problems. In: *Signal Processing* 84 (2004), Aug., S. 957–968

[MPL01] Martin, R. ; Petrovsky, A. ; Lotter, T.: Planar Superdirective Microphone Arrays for Speech Acquisition in the Car. In: *Euro. Conf. Speech Communication and Technology (EUROSPEECH)*. Aalborg, Denmark, Sept. 2001, S. 2623–2626

[MRP96] Mathew, G. ; Reddy, V. U. ; Paulraj, A.: A Quasi-Newton Adaptive Algorithm for Estimating Generalized Eigenvectors. In: *IEEE Transactions on Signal Processing* 44 (1996), Oct., Nr. 10, S. 2413–2422

[MS97] Meyer, J. ; Sydow, C.: Noise Cancelling for Microphone Arrays. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. Munich, Germany, April 1997, S. 211–214

[NA79] Neely, S. T. ; Allen, J. B.: Invertibility of a Room Impulse Response. In: *Journal of the Acoustical Society of America* (1979), July, S. 165–169

[NCB93] Nordholm, S. ; Claesson, I. ; Bengtsson, B.: Adaptive Array Noise Suppression of Handsfree Speaker Input in Cars. In: *IEEE Transactions on Vehicular Technology* 42 (1993), Nov., S. 514–518

[NCG01] Nordholm, S. ; Claesson, I. ; Grbiċ, N.: Optimal and Adaptive Microphone Arrays for Speech Input in Automobiles. In: Brandstein, M.S. (Hrsg.) ; Ward, D.B. (Hrsg.): *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag, 2001, S. 111–132

[NGL05] Nordholm, H. Q. ; Grbic, N. ; Low, S. Y.: Adaptive Microphone Arrays Employing Spatial Quadratic Soft Constraints and Spectral Shaping. In: Benesty, J. (Hrsg.) ; Chen, J. (Hrsg.) ; Makino, S. (Hrsg.): *Speech Enhancement*. Springer-Verlag, 2005, S. 229–246

[NL00] Nordholm, S. ; Leung, Y. H.: Performance Limits of the Broadband Generalized Sidelobe Cancelling Structure in an Isotropic Noise Field. In: *Journal of the Acoustical Society of America* 107 (2000), Feb., S. 1057–1060

[NNS01] Nishiura, T. ; Nakamura, S. ; Shikano, K.: Speech Enhancement by Multiple Beamforming with Reflection Signal Equalization. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. Salt Lake City, USA, May 2001, S. 189–192

[Oja82] Oja, E.: A Simplified Neuron Model as a Principal Component Analyzer. In: *J. Math. Biology* 15 (1982), S. 267–273

[OK85] Oja, E. ; Karhunen, J.: On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. In: *IEEE Transactions Neural Networks* 9 (1985), S. 58–67

[PA02] Parra, L. ; Alvino, C. V.: Geometric Source Separation: Merging Convolutive Source Separation with Geometric Beamforming. In: *IEEE Transactions on Speech and Audio Processing* 10 (2002), Sept., S. 352–362

[PK01] Pados, D. A. ; Karystinos, G. N.: An iterative Algorithm for the Computation of the MVDR Filter. In: *IEEE Transactions on Signal Processing* 49 (2001), Feb., S. 290–300

[QBC88] Quackenbush, S. R. ; Barnwell, T. P. ; Clements, M. A.: *Objective Measures of Speech Quality*. New York : Prentice-Hall, 1988

[RAG04] Ristic, B. ; Arulampalam, S. ; Gordon, N.: *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House Publishers, 2004

[RBB03] Rosca, J. ; Balan, R. ; Beaugeant, C.: Multi-Channel Psychoacoustically Motivated Speech Enhancement. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. HongKong, China, April 2003, S. 84–87

[RBR01] Rickard, S. ; Balan, R. ; Rosca, J.: Real-Time Time-Frequency based Blind Source Separation. In: *Proc. of the second international workshop on independent component analysis and blind signal separation*. San Diego, USA, Dec. 2001, S. 651–656

[RC03] Rindel, J. H. ; Christensen, C. L.: Room Acoustic Simulation and Suralization - How close can we get to the real room. In: *Eight Western Pacific Acoustics conference*. Melbourne, April 2003

[RGC07a] Reuven, G. ; Gannot, S. ; Cohen, I.: Joint Noise Reduction and Acoustic Echo Cancellation using the Transfer-Function Generalized Sidelobe Canceller. In: *Speech Communication - Speech Enhancement* 49 (2007), Aug., S. 623–635

[RGC07b] Reuven, G. ; Gannot, S. ; Cohen, I.: Performance Analysis of Dual Source Transfer-Function Generalized Sidelobe Canceller. In: *Speech Communication - Speech Enhancement* 49 (2007), Aug., S. 623–635

[RGC08] Reuven, G. ; Gannot, S. ; Cohen, I.: Dual-Source Transfer-Function Generalized Sidelobe Canceller. In: *IEEE Transactions on Audio, Speech and Language Processing* 16 (2008), May, Nr. 4

[RHK05] Rohdenburg, T. ; Hohmann, V. ; Kollmeier, B.: Objective Perceptual Quality Measures for the Evaluation of Noise Reduction Schemes. In: *International Workshop on Acoustic Echo and Noise Control*. Eindhoven, Sept. 2005, S. 169–172

[RM05] Rombouts, G. ; Moonen, M.: Fast QRD-Lattice-based unconstrained Optimal Filtering for Acoustic Noise Reduction. In: *IEEE Transactions on Speech and Audio Processing* 13 (2005), Nov., Nr. 6, S. 1130–1143

[RP02] Rao, Y. N. ; Principe, J. C.: Time Series Segmentation Using a Novel Adaptive Eigendecomposition Algorithm. In: *Journal of VLSI Signal Processs* 32 (2002), Nr. 1-3, S. 7–12

[RPW04] Rao, Y. N. ; Principe, J. C. ; Wong, T. F.: Fast RLS-Like Algorithm for Generalized Eigendecomposition and its Applications. In: *Journal of VLSI Signal Processs* 37 (2004), Nr. 2-3, S. 333–344

[RRFM98] Rabinkin, D. ; Renomeron, R. ; Flanagan, J. ; Macomber, D. F.: Optimal Truncation Time for Matched Filter Array Processing. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. Seattle, USA, May 1998, S. 3269–3273

[RSB+05] Ramirez, J. ; Segura, J.C. ; Benitez, C. ; Garcia, L. ; Rubio, A.: Statistical Voice Activity Detection using a Multiple Observation Likelihood Ratio Test. In: *IEEE Signal Processing Letters* 12 (2005), Oct., S. 689– 692

[RYPD05] Raykar, V. C. ; Yegnanarayana, B. ; Prasanna, S. R. M. ; Duraiswami, R.: Source Localization in Reverberant Environments: Modeling and Statistical Analysis. In: *IEEE Transactions on Speech and Audio Processing* 13 (2005), Sept., S. 751–760

[Sab22] Sabine, W. C.: Collected Papers on Acoustics. In: *Harvard University Press, reprinted by Peninsula Publishing, Acous. Soc. Am. 1993 edition* (1922)

[SBM01] Simmer, K. U. ; Bitzer, J. ; Marro, C.: Post-filtering techniques. In: Brandstein, M.S. (Hrsg.) ; Ward, D.B. (Hrsg.): *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag, 2001, S. 39–57

[Sch65] Schroeder, M. R.: New Method of Measuring Reverberation Time. In: *Journal of the Acoustical Society of America* 37 (1965), S. 409–412

[Sch79] Schmidt, R. O.: Multiple Emitter Location and Signal Parameter Estimation. In: *Proc. RADC Spectrum Estimation Workshop*. Rome, NY, USA, 1979, S. 243–258

[SHU06] Schmalenstroeer, J. ; Haeb-Umbach, R.: Online Speaker Change Detection by Combining BIC with Microphone Array Beamforming. In: *Proc. Interspeech*. Pittsburgh, USA, Sept. 2006

[SHU07] Schmalenstroeer, J. ; Haeb-Umbach, R.: Joint Speaker Segmentation, Localization and Identification for Streaming Audio. In: *Proc. Interspeech*. Antwerp, Belgium, Aug. 2007

[SHUW07] Schmalenströer, J. ; Häb-Umbach, R. ; Warsitz, E.: Projekt Amigo - Sprachsignalverarbeitung im vernetzten Haus. In: *Fortschritte der Akustik - DAGA 2007, DEGA e.V.* Stuttgart, März 2007, S. 631–632

[Shy92] Shynk, J.: Frequency-Domain and Multirate Adaptive Filtering. In: *IEEE Signal Processing Magazine* 9 (1992), S. 14–39

[SK06] Schwarz, H.-R. ; Köckler, N.: *Numerische Mathematik*. Teubner, 2006

[SKS99] Sohn, J. ; Kim, N. ; Sung, W.: A Statistical Model-based Voice Activity Detection. In: *IEEE Signal Processing Letters* 6 (1999), Jan., S. 1–3

[SMH+03] Shoko, A. ; Makino, S. ; Hinamoto, Y. ; Mukai, R. ; Nishikawa, T. ; Saruwatari, H.: Equivalence between Frequency-Domain Blind Source Separation and Frequency-Domain Adaptive Beamforming for Convolutive Mixtures. In: *EURASIP Journal on Applied Signal Processing*, 2003, S. 1157–1166

[SMM05] Sawada, H. ; Mukai, S. ; Makino, S.: Frequency-Domain Blind Source Separation. In: Benesty, J. (Hrsg.) ; Chen, J. (Hrsg.) ; Makino, S. (Hrsg.): *Speech Enhancement*. Springer-Verlag, 2005, S. 299–352

[SMW02] Spriet, A. ; Moonen, M. ; Wouters, J.: A multichannel subband gsvd approach to speech enhancement. In: *Eur. Trans. Telecommunications, Special Issue on Acoustic Echo and Noise Control* 13 (2002), March, S. 149–158

[SRS04] Seltzer, M. L. ; Raj, B. ; Stern, R. M.: Likelihood Maximizing Beamforming for Robust Hands-Free Speech Recognition. In: *IEEE Transactions on Speech and Audio Processing* 12 (2004), Sept., S. 489–498

[SSR01] Strobel, N. ; Spors, S. ; Rabenstein, R.: Joint Audio-Video Signal Processing for Object Localization and Tracking. In: Brandstein, M.S. (Hrsg.) ; Ward, D.B. (Hrsg.): *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag, 2001, S. 204–225

[SW92] Simmer, K. U. ; Wasiljeff, A.: Adaptive Microphone Arrays for Noise Suppression in the Frequency Domain. In: *Second Cost 229 Workshop on Adaptive Algorithms in Communications*. Bordeaux, France, Oct. 1992, S. 185–194

[SW96] Shalvi, O. ; Weinstein, E.: System Identification using Nonstationary Signals. In: *IEEE Transactions on Signal Processing* (1996), Aug., S. 2055–2063

[Thi53] Thiele, R.: Richtungsverteilung und Zeitfolge der Schallrückwürfe in Räumen. In: *Acustica 3*, 1953, S. 291–302

[Tuc92] Tucker, R.: Voice Activity Detection Using a Periodicity Measure. In: *IEEE Signal Processing Letters* 139 (1992), Aug., S. 377–380

[TV07] Tran Vu, D. H.: *Akustische Quellentrennung durch adaptives Beamforming basierend auf Verfahren zur Eigenwertzerlegung*. 2007. – Diplomarbeit 4/06, Fachgebiet Nachrichtentechnik, Universität Paderborn

[Täg98] Täger, W.: Near Field Superdirectivity (NFSD). In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. Atlanta, USA, May 1998, S. 2045–2048

[US56] Uzsoky, M. ; Solymar, L.: Theory of super-directive linear arrays. In: *Acta. Physica Hungarica* 6 (1956), May, S. 185–205

[VB01] Vermaak, J. ; Blake, A.: Nonlinear Filtering for Speaker Tracking in Noisy and Reverberant Environments. In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*. Salt Lake City, USA, April 2001

[VHH98] Vary, P. ; Heute, U. ; Hess, W.: *Digitale Sprachsignalverarbeitung*. Stuttgart : Teubner Verlag, 1998

[VM06] Vary, P. ; Martin, R.: *Digital Speech Transmission - Enhancement, Coding & Error Concealment*. John Wiley & Sons, 2006

[VMPG29] Von Mises, R. ; Pollaczek-Geiringer, H.: Praktische Verfahren der Gleichungsauflösung. In: *Zeitschrift für Angewandte Mathematik und Mechanik* (1929), 9, S. 58–79; 152–164

[VSO97] Viberg, M. ; Stoica, P. ; Ottersten, B.: Maximum Likelihood Array Processing in Spatially Correlated Noisefields using Parameterized Signals. In: *IEEE Transactions on Acoustics, Speech and Signal Processing* 45 (1997), April, S. 996–1004

[VT68] Van Trees, H. L.: *Detection, Estimation, and Modulation Theory, Part I.* John Wiley & Sons, 1968

[VT02] Van Trees, H. L.: *Optimum Array Processing.* John Wiley & Sons, 2002

[VVB88] Van Veen, B. D. ; Buckley, K. M.: Beamforming: A Versatile Approach to Spatial Filtering. In: *IEEE Trans. Acoust., Speech, Signal Processing* 5 (1988), Nr. 4, S. 4–24

[WA96] Wax, M ; Anu, Y.: Performance Analysis of the Minimum Variance Beamformer in the Presence of Steering Vector Errors. In: *IEEE Transactions on Signal Processing* 44 (1996), April, S. 938–947

[WB98] Wang, C. ; Brandstein, M. S.: A Hybrid Real-Time Face Tracking System. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP).* Seattle, USA, May 1998, S. 3737–3740

[Wel67] Welch, P.: The use of Fast Fourier Transform for the Estimation of Power Spectra: A Method based on Time Averaging over Short, Modified Periodograms. In: *IEEE Transactions on Audio and Electroacoustics* 15 (1967), June, S. 70–73

[WHU04] Warsitz, E. ; Haeb-Umbach: Robust Speaker Direction Estimation with Particle Filtering. In: *IEEE Workshop on Multimedia Signal Processing (MMSP).* Siena, Italy, Sept. 2004, S. 367– 370

[WHU05] Warsitz, E. ; Haeb-Umbach, R.: Acoustic Filter-and-Sum Beamforming by Adaptive Principal Component Analysis. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP).* Philadelphia, USA, March 2005

[WHU06a] Warsitz, E. ; Haeb-Umbach, R.: Controlling Speech Distortion in Adaptive Frequency-Domain Principal Eigenvector Beamforming. In: *Int. Workshop on Acoustic Echo and Noise Control (IWAENC).* Paris, France, Sept. 2006

[WHU06b] Warsitz, E. ; Haeb-Umbach, R.: Mehrkanalige Sprachsignalverarbeitung durch adaptives Eigenbeamforming für Freisprecheinrichtungen im Kraftfahrzeug. In: *Fortschritte der Akustik, DAGA 2006* Bd. 32. Braunschweig, März 2006, S. 49–50

[WHU07] Warsitz, E. ; Haeb-Umbach, R.: Blind Acoustic Beamforming based on Generalized Eigenvalue Decomposition. In: *IEEE Transactions on Audio, Speech and Language Processing* 15 (2007), July, S. 1529–1539

[WHUP04] Warsitz, E. ; Haeb-Umbach, R. ; Peschke, S.: Adaptive Beamforming Combined with Particle Filtering for Acoustic Source Localization. In: *Proc. ICSLP*. Jeju, Corea, Oct. 2004, S. 2849–2852

[WHUS07] Warsitz, E. ; Häb-Umbach, R. ; Schmalenströer, J.: Zweistufige Sprache/Pause-Detektion in stark gestörter Umgebung. In: *Fortschritte der Akustik - DAGA 2007, DEGA e.V.* Stuttgart, März 2007, S. 303–304

[WHUTV07] Warsitz, E. ; Haeb-Umbach, R. ; Tran Vu, D. H.: Blind Adaptive Principal Eigenvector Beamforming for Acoustical Source Separation. In: *Proc. Interspeech*. Antwerp, Belgium, Aug. 2007

[Wie44] Wielandt, H.: *Beiträge zur mathematischen Behandlung komplexer Eigenwertprobleme*. 1944. – Teil V: Bestimmung höherer Eigenwerte durch gebrochene Iteration. Bericht B 44/J/37, Aerodynamische Versuchsanstalt Göttingen, Germany, 1944

[WKHU08] Warsitz, E. ; Krueger, A. ; Haeb-Umbach, R.: Speech Enhancement with a new Generalized Eigenvector Blocking Matrix for Application in a Generalized Sidelobe Canceller. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. Las Vegas, USA, March/April 2008, S. 73–76

[WKW01] Ward, D. B. ; Kennedy, R. A. ; Williamson, R. C.: Constant Directivity Beamforming. In: Brandstein, M.S. (Hrsg.) ; Ward, D.B. (Hrsg.): *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag, 2001, S. 3–17

[WLW03] Ward, D. B. ; Lehmann, E. A. ; Williamson, R. C.: Particle Filtering Algorithms for Tracking an Acoustic Source in a Reverberant Environment. In: *IEEE Transactions on Speech and Audio Processing* 11 (2003), Nov., S. 826–836

[WMGG67] Widrow, B. ; Mantey, P. E. ; Griffiths, L. J. ; Goode, B. B.: Adaptive Antenna Systems. In: *IEEE Proceedings* 55 (1967), Dec., S. 2143– 2159

[WW02] Ward, D. B. ; Williamson, R. C.: Particle Filter Beamforming for Acoustic Source Location. In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*. Orlando, USA, May 2002

[Yan95] Yang, B.: Projection Approximation Subspace Tracking. In: *IEEE Transactions Signal Processing* 43 (1995), Jan., S. 95–107

[YOZC04] Yang, K. ; Ohira, T. ; Zhang, Y. ; Chi, C.-Y.: Super-Exponential Blind Adaptive Beamforming. In: *IEEE Transactions on Signal Processing* 52 (2004), June, Nr. 6, S. 1549–1563

[YR04] Yilmaz, O. ; Richard, S.: Blind Separation of Speech Mixtures via Time-Frequency Masking. In: *IEEE Transactions on Signal Processing* 52 (2004), July, S. 1830–1847

[YXYZ06] Yang, J. ; Xi, H. ; Yang, F. ; Zhao, Y.: A Quasi-Newton Adaptive Algorithm for Estimating Generalized Eigenvectors. In: *IEEE Transactions on Signal Processing* 44 (2006), Oct., Nr. 10, S. 1177– 1188

[Zel88] Zelinski, R.: A Microphone Array with Adaptive Post-Filtering for Noise Reduction in Reverberant Rooms. In: *Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. New York, USA, April 1988, S. 2578–2581

[ZHA04] Zhang, X. ; Hansen, J. H. L. ; Arehart, K.: Speech Enhancement based on a combined Multi-Channel Array with Constrained Interative and Auditory Masked Processing. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*. Montreal, Canada, May 2004, S. 229–232

Eigene Publikationen

[1] Krueger, A. ; Warsitz, E. ; Haeb-Umbach, R.: Eigenvector based Transfer Function Ratios Estimation for Speech Enhancement with a GSC-like Structure. In: *IEEE Transactions on Audio, Speech and Language Processing*, submitted June 2008

[2] Warsitz, E. ; Krueger, A. ; Haeb-Umbach, R.: Speech Enhancement with a new Generalized Eigenvector Blocking Matrix for Application in a Generalized Sidelobe Canceller. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP).* Las Vegas, USA, March/April 2008, S. 73–76

[3] Häb-Umbach, R. ; Krüger, A. ; Warsitz, E.: Blinde akustische Strahlformung für Anwendungen im KFZ. In: *Fortschritte der Akustik - DAGA 2008, DEGA e.V.* Dresden, März 2008

[4] Warsitz, E. ; Haeb-Umbach, R. ; Tran Vu, D. H.: Blind Adaptive Principal Eigenvector Beamforming for Acoustical Source Separation. In: *Proc. Interspeech.* Antwerp, Belgium, Aug. 2007

[5] Warsitz, E. ; Haeb-Umbach, R.: Blind Acoustic Beamforming based on Generalized Eigenvalue Decomposition. In: *IEEE Transactions on Audio, Speech and Language Processing* 15 (2007), July, S. 1529–1539

[6] Schmalenströer, J. ; Häb-Umbach, R. ; Warsitz, E.: Projekt Amigo - Sprachsignalverarbeitung im vernetzten Haus. In: *Fortschritte der Akustik - DAGA 2007, DEGA e.V.* Stuttgart, März 2007, S. 631–632

[7] Warsitz, E. ; Häb-Umbach, R. ; Schmalenströer, J.: Zweistufige Sprache/Pause-Detektion in stark gestörter Umgebung. In: *Fortschritte der Akustik - DAGA 2007, DEGA e.V.* Stuttgart, März 2007, S. 303–304

[8] Warsitz, E. ; Haeb-Umbach, R.: Controlling Speech Distortion in Adaptive Frequency-Domain Principal Eigenvector Beamforming. In: *Int. Workshop on Acoustic Echo and Noise Control (IWAENC).* Paris, France, Sept. 2006

[9] Warsitz, E. ; Haeb-Umbach, R.: Mehrkanalige Sprachsignalverarbeitung durch adaptives Eigenbeamforming für Freisprecheinrichtungen im Kraftfahrzeug. In: *Fortschritte der Akustik, DAGA 2006* Bd. 32. Braunschweig, März 2006, S. 49–50

[10] Haeb-Umbach, R. ; Warsitz, E.: Adaptive Filter-and-Sum Beamforming in Spatially Correlated Noise. In: *Int. Workshop on Acoustic Echo and Noise Control (IWAENC).* Eindhoven, Netherlands, Sept. 2005

[11] Warsitz, E. ; Haeb-Umbach, R.: Acoustic Filter-and-Sum Beamforming by Adaptive Principal Component Analysis. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP).* Philadelphia, USA, March 2005

[12] Warsitz, E. ; Haeb-Umbach, R. ; Peschke, S.: Adaptive Beamforming Combined with Particle Filtering for Acoustic Source Localization. In: *Proc. ICSLP*. Jeju, Corea, Oct. 2004, S. 2849–2852

[13] Warsitz, E. ; Haeb-Umbach: Robust Speaker Direction Estimation with Particle Filtering. In: *IEEE Workshop on Multimedia Signal Processing (MMSP)*. Siena, Italy, Sept. 2004, S. 367– 370