

Optimierung frequenzvarianter Nullbeamformer für akustische Signale mittels Statistik höherer Ordnung - Anwendungen im Kfz und in Bürräumen

Von der Fakultät Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Verleihung des akademischen Grades
Doktor-Ingenieur
genehmigte Dissertation

vorgelegt von Dipl.-Ing. Wolf Baumann

Promotionsausschuss:

Vorsitzender: Prof. Dr.-Ing. Clemens Gühmann

Gutachter: Prof. Dr.-Ing. Reinhold Orglmeister

Gutachter: Prof. Dr.-Ing. Dieter Filbert

Tag der wissenschaftlichen Aussprache: 14. Dezember 2005

Berlin 2005

D 83

Vorwort

Diese Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Fachgebiet Elektronik und medizinische Signalverarbeitung der Technischen Universität Berlin.

Mein besonderer Dank gilt Herrn Prof. Dr.-Ing. R. Orglmeister für die Unterstützung und die Betreuung dieser Arbeit.

Herrn Prof. Dr.-Ing. D. Filbert danke ich für die Übernahme des Koreferats.

Weiterhin möchte ich mich auch bei den Mitarbeitern und Kollegen am Lehrstuhl für die angenehme Zusammenarbeit und für die fachliche Unterstützung bedanken. Besonderer Dank gilt dabei meinen Kollegen Frau Dipl.-Ing. Dorothea Kolossa und Herrn Dr.-Ing. Bert-Uwe Köhler für die vielen Anregungen und fruchtbaren Diskussionen.

Schließlich gilt herzlicher Dank auch meiner Familie, die mich während der gesamten Zeit tatkräftig unterstützt hat.

Berlin, September 2005

Wolf Baumann.

Inhaltsverzeichnis

| | | |
|----------|--|-----------|
| 1 | Einleitung | 1 |
| 2 | Mehrkanalige Störgeräuschunterdrückung | 3 |
| 2.1 | Prinzipien der mehrkanaligen Störgeräuschunterdrückung | 4 |
| 2.1.1 | Kohärenzfunktion | 4 |
| 2.1.2 | Unterdrückung kohärenter Störsignale | 7 |
| 2.1.3 | Unterdrückung inkohärenter Störsignale | 10 |
| 2.2 | Praktische Gesichtspunkte | 12 |
| 3 | Beamforming-Verfahren | 15 |
| 3.1 | Theoretische Grundlagen | 15 |
| 3.1.1 | Blockweise Verarbeitung | 15 |
| 3.1.2 | Ausbreitung von Schallwellen | 17 |
| 3.1.3 | Mikrofonarrays | 18 |
| 3.2 | Einteilung und Bewertung von Beamformern | 21 |
| 3.2.1 | Einteilung von Beamformern | 21 |
| 3.2.2 | Bewertung von Beamformern | 22 |
| 3.2.2.1 | Beampattern | 22 |
| 3.2.2.2 | Array Gain | 23 |
| 3.2.2.3 | White Noise Gain (WNG) | 24 |
| 3.3 | Nichtadaptive Beamformer | 26 |
| 3.3.1 | Delay-and-Sum-Beamformer | 27 |
| 3.3.2 | Filter-and-Sum-Beamformer | 29 |
| 3.3.2.1 | Superdirektiver Beamformer | 30 |
| 3.3.2.2 | Superdirektiver Entwurf für kohärente Störsignale | 34 |
| 3.4 | Adaptive Beamformer | 39 |
| 3.4.1 | Linear Constrained Minimum Variance Beamformer | 39 |

| | | |
|----------|---|-----------|
| 3.4.2 | Generalized Sidelobe Canceller (GSC) | 41 |
| 4 | Methoden der blinden Quellentrennung | 43 |
| 4.1 | Signalmodelle | 43 |
| 4.1.1 | Modell der linearen, instantanen Überlagerung | 43 |
| 4.1.2 | Modell der linearen, konvolutiven Überlagerung | 45 |
| 4.1.2.1 | Entmischungsstrukturen | 48 |
| 4.2 | Blinde Quellentrennung instantaner Mischungen | 50 |
| 4.2.1 | Definition der statistischen Unabhängigkeit | 50 |
| 4.2.1.1 | Dekorrelation als Vorverarbeitung | 51 |
| 4.2.2 | Prinzipien der Informationstheorie | 52 |
| 4.2.2.1 | Konzept der Minimierung der Einzelentropien | 54 |
| 4.2.2.2 | Konzept der Maximierung der Gesamtentropie | 55 |
| 4.2.3 | ICA durch Maximum-Likelihood-Schätzung | 57 |
| 4.2.4 | Kumulantenbasierte ICA-Verfahren | 59 |
| 4.2.5 | Verfahren auf der Basis von Statistik zweiter Ordnung | 61 |
| 4.3 | Blinde Quellentrennung konvolutiver Mischungen | 62 |
| 4.3.1 | Kombination von ICA und Beamforming | 65 |
| 4.3.1.1 | Beampattern für Permutationskorrektur | 65 |
| 5 | Optimierung frequenzvarianter Nullbeamformer mittels Statistik höherer Ordnung | 69 |
| 5.1 | Vergleich von Beamforming und konvolutiver ICA | 69 |
| 5.1.1 | Parametervergleich | 69 |
| 5.1.2 | Strukturvergleich | 71 |
| 5.1.3 | Vergleich der Kostenfunktionen | 71 |
| 5.2 | Frequenzvarianter Nullbeamformer mit ICA-Kostenfunktion | 72 |
| 5.2.1 | Motivation | 72 |
| 5.2.2 | Funktionsprinzip | 73 |
| 5.2.3 | Kostenfunktion | 75 |
| 5.2.4 | Optimierung | 78 |
| 5.2.5 | Praktische Gesichtspunkte | 82 |
| 5.2.5.1 | Realzeitfähigkeit | 82 |
| 5.2.5.2 | Raumbedingungen | 84 |
| 5.2.5.3 | Mikrofonabstand | 86 |
| 5.2.5.4 | Erweiterbarkeit auf mehrere Quellen | 90 |
| 5.2.5.5 | Berücksichtigung des Betragsanteils | 90 |

| | | |
|-----------------------------|---|------------|
| 6 | Ergebnisse und Diskussion | 93 |
| 6.1 | Qualitätsmaße | 93 |
| 6.1.1 | Störsignaldämpfung | 94 |
| 6.1.2 | Nutzsignalstörung | 95 |
| 6.2 | Datensätze und Meßbedingungen | 97 |
| 6.2.1 | Erzeugung der synthetischen Mischungen | 97 |
| 6.2.2 | Aufnahme realer Mischungen | 97 |
| 6.2.2.1 | Aufnahmen im Laborraum | 97 |
| 6.2.2.2 | Aufnahmen im Automobil | 98 |
| 6.3 | Ergebnisse der synthetischen Mischungen | 99 |
| 6.3.1 | Ergebnistabellen der synthetischen Mischungen | 101 |
| 6.4 | Ergebnisse der realen Aufnahmen | 105 |
| 6.5 | Diskussion | 106 |
| 7 | Zusammenfassung | 111 |
| A | Gradientenberechnung | 113 |
| B | Reduktion des Rechenaufwandes | 117 |
| B.1 | Aufwandsreduzierte Kostenfunktion | 117 |
| B.2 | Aufwandsreduzierter Gradient | 120 |
| Literaturverzeichnis | | 123 |

Abbildungsverzeichnis

| | | |
|------|--|----|
| 2.1 | Rückspiegel für LINGUATRONIC System von Mercedes-Benz [104] | 3 |
| 2.2 | Modell der mehrkanaligen Signalerfassung | 4 |
| 2.3 | Adaptives Filter mit Störreferenz | 7 |
| 2.4 | Störsignalunterdrückung in Abhängigkeit von der Kohärenz | 9 |
| 2.5 | Störsignalunterdrückung bei unkorreliertem Rauschen | 12 |
| 3.1 | Äquidistantes Linienarray zur mehrkanaligen Signalerfassung | 19 |
| 3.2 | Beampattern bei vier Koeffizienten mit einer konstanten Gewichtung von $\frac{1}{4}$ und einem Mikrofonabstand von $d = 5$ cm | 23 |
| 3.3 | Beampattern bei vier Koeffizienten mit einer konstanten Gewichtung von $\frac{1}{4}$ und einem Mikrofonabstand von $d = 5$ cm in dreidimensionaler Darstellung | 24 |
| 3.4 | Beampattern bei vier Koeffizienten mit einer konstanten Gewichtung von $\frac{1}{4}$ und einem Mikrofonabstand von $d = 5$ cm als Überlagerung aller Frequenzbänder dargestellt | 25 |
| 3.5 | Zweikanaliger Beamformer mit additivem Sensorrauschen | 25 |
| 3.6 | Allgemeine Struktur des Delay-and-Sum-Beamformers | 27 |
| 3.7 | Allgemeine Struktur eines Filter and Sum Beamformers | 29 |
| 3.8 | Superdirektiver Beamfomer bei Nutzsignal aus 20° und vier Mikrofonen mit einem äquidistanten Abstand von $d = 15$ mm | 32 |
| 3.9 | Superdirektiver Beamfomer bei Nutzsignal aus 20° und vier Mikrofonen mit einem äquidistanten Abstand von $d = 15$ mm bei Entwurf mit begrenzter Koeffizientennorm von 10 dB | 33 |
| 3.10 | Norm der Koeffizienten des superdirektiven Beamfomers bei Begrenzung auf 10 dB | 34 |
| 3.11 | Beampattern eines superdirektiven Breitbandbeamformers im Frequenzbereich für vier Mikrofone mit einem äquidistanten Abstand von $d = 15$ mm | 36 |
| 3.12 | Norm der Koeffizienten des superdirektiven Beamformers aus Abbildung 3.11 | 37 |

| | | |
|------|---|-----|
| 3.13 | Beampattern des auf $\mathbf{w}^H \mathbf{w} \leq 6$ dB beschränkten superdirektiven Beamformers mit vier Mikrofonen im äquidistanten Abstand von $d = 15$ mm . . . | 38 |
| 3.14 | Norm der Koeffizienten des regularisierten superdirektiven Beamformers mit vier Mikrofonen im äquidistanten Abstand von $d = 15$ mm | 39 |
| 3.15 | Struktur des Frost-Beamformers | 40 |
| 3.16 | Struktur des Generalized Sidelobe Cancellers | 42 |
| 4.1 | Modell der instantanen Mischung zweier Signale | 44 |
| 4.2 | Modell der konvolutiven Mischung zweier Signale | 46 |
| 4.3 | Modell der normierten, konvolutiven Mischung zweier Signale | 47 |
| 4.4 | Feed-Back-Struktur zur Entmischung | 48 |
| 4.5 | Konnektionistisches Neuron | 55 |
| 4.6 | Funktionsblöcke für frequenzbereichsbasierte ICA-Verfahren | 63 |
| 4.7 | Beispiel für Beampattern von ICA-Entmischungsfiltren bei einem Mikrofonabstand von $d = 3$ cm und Quellsignalen aus 20° und -40° | 66 |
| 4.8 | Gegenüberstellung der Minima der beiden Beampattern aus Abbildung 4.7 | 67 |
| 4.9 | Gegenüberstellung der Permutationskorrektur durch Auswertung der Minima der Beampattern (oben) sowie durch Maximum-Likelihood-Zuordnung nach [7] (unten) | 68 |
| 5.1 | Konvolute Quellentrennung als Parrallelschaltung von zwei Filter-And-Sum-Beamformern | 71 |
| 5.2 | Projektion des Gradientenvektors auf den Einheitskreis | 80 |
| 5.3 | Schema des Optimierungsalgorithmus | 82 |
| 5.4 | Simulierter Raum | 84 |
| 5.5 | Streuung der Einfallswinkel über der Frequenz bei verschiedenen Raumkonfigurationen | 85 |
| 5.6 | Streuung der Einfallsrichtung in Abhängigkeit vom Mikrofonabstand . . . | 87 |
| 5.7 | Winkel der Phasenfaktoren bei Mikrofonabstand $d=1$ cm und verschiedenen Frequenzen | 88 |
| 5.8 | Rauschverstärkung des zweikanaligen Nullbeamformers bei Mikrofonabstand $d=1$ cm | 89 |
| 5.9 | Amplitudenunterschiede in Abhängigkeit vom Mikrofonabstand | 91 |
| 6.1 | Messanordnung im Laborraum | 98 |
| 6.2 | Positionen der Sprecher im Auto | 99 |
| 6.3 | Beispiel für die Abhängigkeit der Rechenzeit von der Nachhallzeit bei synthetischen Mischungen | 107 |

| | | |
|-----|---|-----|
| 6.4 | Abhängigkeit der Trennungsleistung von der Nachhallzeit | 108 |
| 6.5 | Abhängigkeit der Trennungsleistung von der FFT-Länge | 109 |

Tabellenverzeichnis

| | | |
|------|--|-----|
| 5.1 | Anzahl der freien Parameter von Beamforming und blinder Quellentrennung | 70 |
| 6.1 | Übersicht über die Konfigurationen der Messungen im Laborraum | 98 |
| 6.2 | Konfigurationen der Messungen im Auto | 99 |
| 6.3 | Übersicht über die Ergebnistabellen der synthetischen Mischung | 100 |
| 6.4 | Kurzbezeichnung der bei synthetischen Mischungen verwendeten Verfahren | 100 |
| 6.5 | Ergebnisse für NBF , Sprecherkombination Ii/Fm, d=2.5cm in Abhängigkeit von der Nachhallzeit | 101 |
| 6.6 | Ergebnisse für NBFb , Sprecherkombination Ii/Fm, d=2.5cm in Abhängigkeit von der Nachhallzeit | 101 |
| 6.7 | Ergebnisse für NBF , Sprecherkombination Ii/Fm, d=5cm in Abhängigkeit von der Nachhallzeit | 102 |
| 6.8 | Ergebnisse für NBFb , Sprecherkombination Ii/Fm, d=5cm in Abhängigkeit von der Nachhallzeit | 102 |
| 6.9 | Ergebnisse für NBF , Sprecherkombination Ar/Ga, d=2.5cm in Abhängigkeit von der Nachhallzeit | 103 |
| 6.10 | Ergebnisse für NBFb , Sprecherkombination Ar/Ga, d=2.5cm in Abhängigkeit von der Nachhallzeit | 103 |
| 6.11 | Ergebnisse für NBF , Sprecherkombination Ar/Ga, d=5cm in Abhängigkeit von der Nachhallzeit | 104 |
| 6.12 | Ergebnisse für NBFb , Sprecherkombination Ar/Ga, d=5cm in Abhängigkeit von der Nachhallzeit | 104 |
| 6.13 | Kurzbezeichnung der bei realen Aufnahmen verwendeten Verfahren | 105 |
| 6.14 | Ergebnisse im Laborraum bei Konfiguration A | 106 |
| 6.15 | Ergebnisse im Laborraum bei Konfiguration B | 106 |
| 6.16 | Ergebnisse im Kfz bei Konfiguration A | 106 |
| 6.17 | Ergebnisse im Kfz bei Konfiguration B | 107 |

6.18 Ergebnisse im Kfz bei Konfiguration C 107

Notation

| Bezeichnung | Erläuterung |
|----------------------------|---------------------------------|
| \mathbf{A} | Mischungsmatrix |
| c | Ausbreitungsgeschwindigkeit |
| \mathbf{C} | Kovarianzmatrix |
| cum | Kumulante |
| d | Mikrofonabstand |
| $D(p_x p_y)$ | Kullback-Leibler-Divergenz |
| $E[\cdot]$ | Erwartungswert |
| $e_i(\omega, \varphi)$ | Phasenfaktoren |
| f | Frequenz |
| $\gamma_{x_1 x_2}(\omega)$ | Kohärenzfunktion |
| $I(x_1, \dots, x_N)$ | gemeinsame Information |
| \mathbf{I} | Einheitsmatrix |
| $\Im(\cdot)$ | Imaginärteil |
| \mathbf{k} | Wellenzahlvektor |
| κ | Kurtosis |
| \mathcal{K} | Kostenfunktion |
| λ | Wellenlänge |
| L | Blocklänge der DFT |
| $n(t)$ | Rauschsignal im Zeitbereich |
| $N(\omega)$ | Rauschsignal im Frequenzbereich |
| \mathcal{N} | Negentropie |
| ρ | Reflexionskoeffizient |
| $\mathbf{r}(\omega, \phi)$ | Steering-Vektor |
| $\Re(\cdot)$ | Realteil |
| $r_{x_i x_i}(\tau)$ | Autokorrelationsfunktion |
| $r_{x_i x_j}(\tau)$ | Kreuzkorrelationsfunktion |
| $s_i(t)$ | Quellsignal i im Zeitbereich |

| | |
|------------------------------|--|
| $S_i(\omega)$ | Quellsignal i im Frequenzbereich |
| σ^2 | Varianz |
| SIR_i^+ | Störsignalunterdrückung als SIR-Verbesserung |
| $\Delta\mathcal{S}^{(i)}$ | Maß der Nutzsinalverzerrung |
| \mathbf{w} | Koeffizientenvektor |
| \mathbf{W} | Entmischungsmatrix |
| $\varphi(\omega)$ | frequenzvariante Einfallrichtung |
| $\Phi_{x_i x_i}(\omega)$ | Autoleistungsdichtespektrum |
| $\Phi_{x_i x_j}(\omega)$ | Kreuzleistungsdichtespektrum |
| $\Phi_{\mathbf{xx}}(\omega)$ | Matrix der Sensor-Kreuzleistungsdichtespektren |
| $\Psi(\omega, \phi)$ | Beampattern |
| $x_i(t)$ | Sensorsignal i im Zeitbereich |
| $X_i(\omega)$ | Sensorsignal i im Frequenzbereich |
| $y_i(t)$ | Ausgangssignal i im Zeitbereich |
| $Y_i(\omega)$ | Ausgangssignal i im Frequenzbereich |

Abkürzungen

| Abkürzung | Bedeutung |
|-----------|--------------------------------|
| AKF | Autokorrelationsfunktion |
| BSS | Blind Source Separation |
| DFT | Diskrete Fouriertransformation |
| GSC | Generalized Sidelobe Canceller |
| ICA | Independent Component Analysis |
| KKF | Kreuzkorrelationsfunktion |
| MSC | Magnitude Squared Coherence |
| NBF | Nullbeamformer |
| PCA | Principal Component Analysis |
| SIR | Signal to Interference Ratio |
| SNR | Signal to Noise Ratio |
| STFT | Short Time Fourier Transform |
| WNG | White Noise Gain |

1 Einleitung

Im Zuge der fortschreitenden technologischen Entwicklung erlangen elektronische Kommunikationssysteme eine immer größere Bedeutung. So ist in den letzten Jahren ein stetiges Wachsen der Anzahl sprachgesteuerter Anwendungen zu verzeichnen, und die mobile Kommunikation hat Einzug in das tägliche Leben gehalten. In diesem Zusammenhang wächst gleichzeitig der Bedarf an praxistauglichen Systemen zur Störgeräuschunterdrückung, um in den verschiedenen akustischen Umgebungen eine hohe Sprachverständlichkeit zu gewährleisten.

Aufgrund der gesunkenen Kosten für leistungsfähige Hardware lassen sich zu diesem Zweck mittlerweile auch komplexere Verfahren, wie z.B. adaptive mehrkanalige Algorithmen umsetzen. Solche Verfahren für *Mikrofonarrays* sind einkanaligen Methoden oft überlegen, da sie nicht nur zeitliche, sondern auch räumliche Eigenschaften der zu verarbeitenden Signale ausnutzen können. Sie werden klassischer Weise als *Beamformer* bezeichnet.

Adaptive Beamformer sind dabei auf die Vorgabe der Einfallrichtung des Nutzsignals angewiesen, während die Richtungen der Störsignale in Abhängigkeit von der akustischen Situation erlernt bzw. nachgeregelt werden. In realen akustischen Umgebungen ist die Vorgabe einer einzelnen, festen Nutzsinalrichtung jedoch unrealistisch, da diese durch die mehrfachen Reflexionen innerhalb eines Raumes verändert wird. In Folge dessen kann es zu einer erheblichen Verzerrung des Nutzsignals und damit zur Verschlechterung der Sprachqualität kommen.

Dieses Problem kann methodisch nur durch die zusätzliche Adaption der Nutzsinalcharakteristik bzw. -richtung umgangen werden. Ein dafür geeignetes Hilfsmittel ist durch den Einsatz von Statistik höherer Ordnung gegeben. Insbesondere die Analyse der statistischen Unabhängigkeit (*ICA Independent Component Analysis*) stellt eine probate Methode zur Identifizierung einzelner Komponenten eines Signalgemisches dar. Sie wird aus diesem Grund vor allem für die Lösung von Problemen der blinden Quellentrennung eingesetzt, wo Signalgemische ohne Kenntnis einzelner Anteile in ihre ursprünglichen Komponenten zerlegt werden sollen.

In der vorliegenden Arbeit wird ein Verfahren entwickelt, das die Interpretation der intersensorischen Signalabhängigkeiten als räumliche Parameter ausnutzt, um eine gute Stör-signalunterdrückung zu erreichen. Dabei werden mit Hilfe von Statistik höherer Ordnung direkt die räumlichen Ausbreitungsrichtungen der Signale angepaßt, so daß die explizite Vorgabe einer festen Richtung entfallen kann. Dieses Verfahren läßt sich somit als Kombination der beiden Ansätze *Beamforming* und *Independent Component Analysis* begreifen.

Die Arbeit ist folgendermaßen gegliedert: Im nachfolgenden zweiten Kapitel werden zunächst die nach der Kohärenz der Signale zu unterscheidenden Prinzipien der mehrkanaligen Störgeräuschunterdrückung erläutert, um einen Einblick in die verschiedenen Wirkungsweisen zu geben, und anschließend werden die praktischen Gesichtspunkte bei der Umsetzung der theoretischen Erkenntnisse erläutert. Damit ist die Grundlage für das Verständnis der mehrkanaligen Signalverarbeitungsmethoden *Beamforming* und *blinde Quellentrennung* gelegt, die in Kapitel drei und vier eingehend erläutert werden.

In Kapitel fünf schließt sich ein Vergleich dieser beiden Methoden an, aus dem sich Rückschlüsse auf mögliche Verbesserungen ziehen lassen. Daraus wird ein kombiniertes Verfahren, das als frequenzvariantes Nullbeamforming bezeichnet werden kann, entwickelt. Aufgrund der Struktur und insbesondere der vorteilhaften Parametrisierung erreicht dieses Verfahren eine robuste Unterdrückung kohärenter Störer in Realzeit. Abschließend werden in diesem Kapitel die Möglichkeiten der Verbesserung von Verfahren der blinden Quellentrennung durch die Ausnutzung räumlicher Informationen beleuchtet.

In Kapitel sechs erfolgt die Auswertung der erreichbaren Störsignaldämpfung zum einen durch die Trennung simulierter Überlagerungen und zum anderen anhand realer Raum-aufnahmen. Die Ergebnisse werden im Zusammenhang mit den vorherigen theoretischen Überlegungen diskutiert.

Kapitel sieben faßt die gewonnenen Erkenntnisse zusammen.

2 Mehrkanalige Störgeräuschunterdrückung

Der Begriff *Beamforming* ist eng mit der mehrkanaligen Störgeräuschunterdrückung verbunden. Er stammt ursprünglich von Filtern für mehrere Sensoren, die schmale Strahlengänge zur Unterscheidung räumlich getrennter Signalquellen ausgeformt haben [97]. Dabei wird der Beam, auch als Hauptkeule bezeichnet, auf die interessierende Quelle ausgerichtet, während andere Signaleinfallsrichtungen im Vergleich dazu gedämpft werden.

Die potentiellen Anwendungen dieser Verfahren waren in den frühen 60-er Jahren hauptsächlich in den Gebieten der Sonar- und Radartechnik sowie der seismischen Signalverarbeitung zu finden [29]. Allerdings gab es bereits während des ersten Weltkrieges Versuche, mit Hilfe von Mikrofonarrays feindliche Flugzeuge zu orten [49].

Aufgrund der gesunkenen Kosten für entsprechende Hardwarekomponenten lassen sich heutzutage auch komplexere Beamforming-Verfahren für Mikrofonarrays in der Praxis umsetzen. So ist der Trend zu mehrkanaliger Signalverarbeitung unter anderem im Automobilbereich zu verzeichnen, wo für Freisprecheinrichtungen z.B. vier Mikrofone im Rückspiegel integriert werden (siehe Abbildung 2.1).



Abbildung 2.1: Rückspiegel für LINGUATRONIC System von Mercedes-Benz [104]

2.1 Prinzipien der mehrkanaligen Störgeräuschunterdrückung

Während einkanalige Verfahren der Geräuschreduktion fast immer auf a-priori-Wissen über Nutz- und Störsignal angewiesen sind, lassen sich mit mehrkanaligen Verfahren zusätzlich auch räumliche Informationen auswerten. Damit haben mehrkanalige Verfahren das Potential, bei der Störgeräuschunterdrückung erfolgreicher zu sein.

Um die Prinzipien der mehrkanaligen Störgeräuschunterdrückung näher zu spezifizieren, muß zwischen korrelierten und unkorrelierten Signalen unterschieden werden, die jeweils den Einsatz verschiedener Methoden erfordern. Während korrelierte Signale mit mehrkanaliger Signalverarbeitung im Idealfall vollständig ausgelöscht werden können, ist die Dämpfung für unkorrelierte Signale begrenzt. Die Ursachen dafür sollen nachfolgend auf der Grundlage des in Abbildung 2.2 dargestellten Modells der mehrkanaligen Signalerfassung näher untersucht werden.

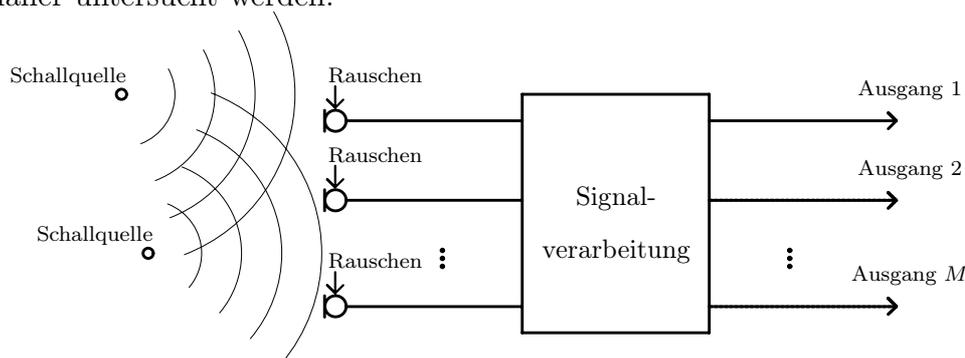


Abbildung 2.2: Modell der mehrkanaligen Signalerfassung

2.1.1 Kohärenzfunktion

Die komplexe Kohärenzfunktion $\gamma(\omega)$ ist ein Maß für den Grad der Korrelation zwischen zwei Signalen und ist für stationäre stochastische Prozesse definiert als das Kreuzleistungsdichtespektrum geteilt durch die Wurzel aus dem Produkt der Autoleistungsdichtespektren [35]. Das Autoleistungsdichtespektrum $\Phi_{x_i x_i}(\omega)$ eines Signals $x_i(t)$ berechnet sich als Fouriertransformierte der Autokorrelationsfunktion

$$r_{x_i x_i}(\tau) = E[x_i(t + \tau)x_i^*(t)] \quad (2.1)$$

wie folgt

$$\Phi_{x_i x_i}(\omega) = \int_{-\infty}^{\infty} r_{x_i x_i}(\tau) e^{-j\omega\tau} d\tau. \quad (2.2)$$

Äquivalent dazu kann das Kreuzleistungsdichtespektrum $\Phi_{x_i x_j}(\omega)$ als Fouriertransformierte der Kreuzkorrelationsfunktion $r_{x_i x_j}$ bestimmt werden

$$\Phi_{x_i x_j}(\omega) = \int_{-\infty}^{\infty} r_{x_i x_j} e^{-j\omega\tau} d\tau. \quad (2.3)$$

Mit diesen Definitionen gilt für die komplexwertige Kohärenzfunktion zweier Signale $x_1(t)$ und $x_2(t)$

$$\gamma_{x_1 x_2}(\omega) = \frac{\Phi_{x_1 x_2}(\omega)}{\sqrt{\Phi_{x_1 x_1}(\omega)\Phi_{x_2 x_2}(\omega)}}. \quad (2.4)$$

Sehr häufig wird auch das reellwertige und zwischen Null und Eins begrenzte Betragsquadrat der Kohärenz, die *Magnitude Squared Coherenc* (MSC), verwendet

$$|\gamma_{x_1 x_2}(\omega)|^2 = \frac{|\Phi_{x_1 x_2}(\omega)|^2}{\Phi_{x_1 x_1}(\omega)\Phi_{x_2 x_2}(\omega)}. \quad (2.5)$$

Da für verschiedene Rauschumgebungen bereits entsprechende Kohärenzfunktionen bekannt sind, können diese beim Entwurf und der Evaluierung von Beamforming-Algorithmen eingesetzt werden. Von besonderer Bedeutung ist dabei das diffuse Rauschfeld, das bei unendlich vielen im Raum verteilten, unabhängigen Schallquellen entsteht [32]. Es weist für zwei im Abstand d platzierte Mikrofone mit Kugelcharakteristik einen si-förmigen Verlauf auf [95]. Mit c als Schallgeschwindigkeit ergibt sich die MSC zu

$$|\gamma_{x_1 x_2}(\omega)|^2 = \left[\frac{\sin(2\pi f \frac{d}{c})}{2\pi f \frac{d}{c}} \right]^2 = \text{si}^2\left(2\pi f \frac{d}{c}\right). \quad (2.6)$$

Bei nur einer aktiven Signalquelle und absoluter Rauschfreiheit beinhalten die beiden Mikrofonensignale $x_1(t)$ und $x_2(t)$ unterschiedlich gefilterte Versionen des aktiven Quellsignals $s(t)$. Mit h_1 und h_2 als Raumimpulsantwort zwischen Schallquelle und jeweiligem Mikrofon ergibt sich im Zeitbereich

$$x_1(t) = h_1(t) * s(t) \quad (2.7)$$

$$x_2(t) = h_2(t) * s(t). \quad (2.8)$$

Bei der Formulierung im Frequenzbereich ist die jeweilige Raumimpulsantwort $h_i(t)$ durch die komplexwertige Übertragungsfunktion $H_i(\omega)$ zu ersetzen, und die Faltung in (2.7) bzw. (2.8) kann als Multiplikation dargestellt werden

$$X_1(\omega) = H_1(\omega) \cdot S(\omega) \quad (2.9)$$

$$X_2(\omega) = H_2(\omega) \cdot S(\omega). \quad (2.10)$$

Für die MSC folgt dann:

$$|\gamma_{x_1x_2}(\omega)|^2 = \frac{|H_1H_2^*\Phi_{ss}|^2}{|H_1|^2\Phi_{ss} \cdot |H_2|^2\Phi_{ss}} \quad (2.11)$$

$$= \frac{|H_1|^2|H_2|^2\Phi_{ss}^2}{|H_1|^2|H_2|^2\Phi_{ss}^2} \quad (2.12)$$

$$= 1. \quad (2.13)$$

Wenn zusätzlich unkorreliertes Rauschen an den Sensoren auftritt und die Mikrofonssignale im Zeitbereich als

$$x_1(t) = h_1 * s(t) + n_1(t) \quad (2.14)$$

$$x_2(t) = h_2 * s(t) + n_2(t) \quad (2.15)$$

bzw. im Frequenzbereich als

$$X_1(\omega) = H_1(\omega)S(\omega) + N_1(\omega) \quad (2.16)$$

$$X_2(\omega) = H_2(\omega)S(\omega) + N_2(\omega) \quad (2.17)$$

beschrieben werden können, verändert sich der konstante Wert aus Gleichung (2.13) zu

$$|\gamma_{x_1x_2}(\omega)|^2 = \frac{|H_1H_2^*\Phi_{ss}|^2}{(|H_1|^2\Phi_{ss} + \Phi_{n_1n_1}) \cdot (|H_2|^2\Phi_{ss} + \Phi_{n_2n_2})}. \quad (2.18)$$

Das Betragsquadrat der Kohärenzfunktion wird also durch das Auftreten von unkorreliertem Rauschen stets verkleinert. Aus diesem Grund läßt sich die Kohärenzfunktion auch als Maß zur Rauschreduktion einsetzen (siehe z.B. [16, 86]).

Bei Rauschfreiheit und zwei statistisch unabhängigen, gerichteten Quellsignalen $s_1(t)$ und $s_2(t)$ nimmt das Modell der Mikrofonssignale die Form

$$x_1(t) = h_1 * s_1(t) + h_2 * s_2(t) \quad (2.19)$$

$$x_2(t) = h_3 * s_1(t) + h_4 * s_2(t) \quad (2.20)$$

an. Analog dazu läßt sich die Formulierung im Frequenzbereich angeben als

$$X_1(\omega) = H_1(\omega)S_1(\omega) + H_2(\omega)S_2(\omega) \quad (2.21)$$

$$X_2(\omega) = H_3(\omega)S_1(\omega) + H_4(\omega)S_2(\omega). \quad (2.22)$$

Für die MSC ergibt sich bei dieser Konfiguration

$$\begin{aligned}
 |\gamma_{x_1 x_2}(\omega)|^2 &= \frac{|H_1 H_3^* \Phi_{s_1 s_1} + H_2 H_4^* \Phi_{s_2 s_2}|^2}{(|H_1|^2 \Phi_{s_1 s_1} + |H_2|^2 \Phi_{s_2 s_2}) \cdot (|H_3|^2 \Phi_{s_1 s_1} + |H_4|^2 \Phi_{s_2 s_2})} \\
 &= \frac{|H_1 H_3|^2 \Phi_{s_1 s_1}^2 + [H_1 H_2^* H_3^* H_4 + H_1^* H_2 H_3 H_4^*] \Phi_{s_1 s_1} \Phi_{s_2 s_2} + |H_2 H_4|^2 \Phi_{s_2 s_2}^2}{|H_1 H_3|^2 \Phi_{s_1 s_1}^2 + [|H_1|^2 |H_4|^2 + |H_2|^2 |H_3|^2] \Phi_{s_1 s_1} \Phi_{s_2 s_2} + |H_2 H_4|^2 \Phi_{s_2 s_2}^2}.
 \end{aligned} \tag{2.23}$$

Dabei unterscheiden sich Nenner und Zähler lediglich in den mit $\Phi_{s_1 s_1} \Phi_{s_2 s_2}$ multiplizierten Termen $[H_1 H_2^* H_3^* H_4 + H_1^* H_2 H_3 H_4^*]$ und $[|H_1|^2 |H_4|^2 + |H_2|^2 |H_3|^2]$. Für diese beiden Terme gilt die Ungleichung

$$[H_1 H_2^* H_3^* H_4 + H_1^* H_2 H_3 H_4^*] \leq [|H_1|^2 |H_4|^2 + |H_2|^2 |H_3|^2]. \tag{2.24}$$

Folglich nimmt das Betragsquadrat der Kohärenzfunktion auch bei Anwesenheit mehrerer gerichteter Signalquellen in den meisten Fällen Werte kleiner Eins an. Dies ist auf die Tatsache zurückzuführen, daß die jeweilige Quelle mit unterschiedlichen komplexen Übertragungsfaktoren in die Mikrofone einkoppelt, die Gleichheit in (2.24) also nur bei identischen Übertragungsfunktionen gegeben wäre.

2.1.2 Unterdrückung kohärenter Störsignale

Als erläuterndes Beispiel für die Unterdrückung kohärenter Störsignale sei das adaptive Filter nach Abbildung 2.3 betrachtet. Es soll zur Rauschunterdrückung für Sprachsignale eingesetzt werden, wobei das Meßsignal $d(t)$, bestehend aus Quellsignal $s(t)$ und Rauschsignal $n(t)$, und ein Referenzrauschsignal $\tilde{n}(t)$ zur Verfügung stehen.

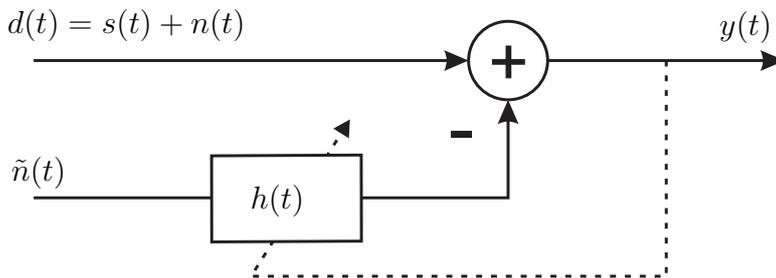


Abbildung 2.3: Adaptives Filter mit Störreferenz

Das optimale Filter im Sinne des kleinsten mittleren quadratischen Fehlers ist das Wienerfilter $\hat{H}_{opt}(\omega) = \frac{\Phi_{d\tilde{n}}(\omega)}{\Phi_{\tilde{n}\tilde{n}}(\omega)}$ [35]. Analysiert man das Verhalten des optimal adaptierten Filters

während einer Sprachpause, also bei $s(t) = 0$, läßt sich das Ausgangs-Leistungsdichtespektrum in Abhängigkeit von den Autoleistungsdichtespektren von Meß- und Referenzsignal wie folgt angeben:

$$\Phi_{yy}(\omega) = \Phi_{dd}(\omega) - H^* \Phi_{d\tilde{n}} - H \Phi_{d\tilde{n}}^* + |\hat{H}_{opt}(\omega)|^2 \Phi_{\tilde{n}\tilde{n}}(\omega). \quad (2.25)$$

Durch Einsetzen von $\hat{H}_{opt}(\omega)$ erhält man

$$\Phi_{yy}(\omega) = \Phi_{dd}(\omega) - \frac{|\Phi_{d\tilde{n}}(\omega)|^2}{\Phi_{\tilde{n}\tilde{n}}(\omega)} - \frac{|\Phi_{d\tilde{n}}(\omega)|^2}{\Phi_{\tilde{n}\tilde{n}}(\omega)} + \frac{|\Phi_{d\tilde{n}}(\omega)|^2}{\Phi_{\tilde{n}\tilde{n}}(\omega)} \quad (2.26)$$

$$= \Phi_{dd}(\omega) - \frac{|\Phi_{d\tilde{n}}(\omega)|^2}{\Phi_{\tilde{n}\tilde{n}}(\omega)} \quad (2.27)$$

$$= \Phi_{dd}(\omega) \left[1 - \frac{|\Phi_{d\tilde{n}}(\omega)|^2}{\Phi_{\tilde{n}\tilde{n}}(\omega) \Phi_{dd}(\omega)} \right] \quad (2.28)$$

$$= \Phi_{dd}(\omega) \left[1 - |\gamma_{d\tilde{n}}|^2 \right]. \quad (2.29)$$

Das Verhältnis von Meßsignalleistung zur Ausgangssignalleistung gibt den Grad der Rauschbefreiung, nachfolgend mit $R(\omega)$ bezeichnet, an

$$R(\omega) = \frac{\Phi_{dd}(\omega)}{\Phi_{yy}(\omega)} \quad (2.30)$$

$$= \frac{\Phi_{dd}(\omega)}{\Phi_{dd}(\omega) \left[1 - |\gamma_{d\tilde{n}}|^2 \right]} \quad (2.31)$$

$$= \frac{1}{1 - |\gamma_{d\tilde{n}}|^2}. \quad (2.32)$$

Gleichung (2.32) beschreibt die erreichbare Störsignaldämpfung in Abhängigkeit von der Kohärenz [87]. Es wird ersichtlich, daß für kohärente Signale im Idealfall eine unendliche Dämpfung bzw. eine Auslöschung erreicht werden kann. Zur zahlenmäßigen Verdeutlichung ist in Abbildung 2.4 auf der nächsten Seite der Verlauf von $R(\omega)$ dargestellt.

Die alleinige Bewertung der MSC als Indikator für die erreichbare Dämpfung ist jedoch nicht ausreichend, wie nachfolgend gezeigt werden soll. Dafür sei ein Beispiel entsprechend dem Modell aus Gleichung (2.7) und (2.8) mit zwei aktiven Schallquellen und den vereinfachten Filtern $h_1 = 1$, $h_2 = 0.1$, $h_3 = 2$ und $h_4 = 3$ betrachtet. Die Mikrofonsignale sind in diesem Fall gegeben als

$$x_1(t) = 1 \cdot s_1(t) + 0.1 \cdot s_2(t) \quad (2.33)$$

$$x_2(t) = 2 \cdot s_1(t) + 3 \cdot s_2(t). \quad (2.34)$$

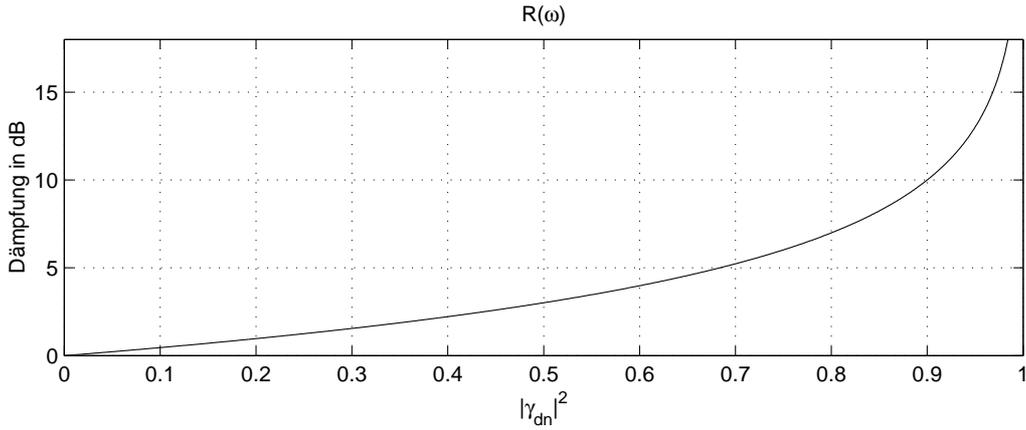


Abbildung 2.4: Störsignalunterdrückung in Abhängigkeit von der Kohärenz

Die MSC berechnet sich entsprechend Gleichung (2.23) zu

$$\begin{aligned}
 |\gamma_{x_1x_2}(\omega)|^2 &= \frac{|H_1H_3^*\Phi_{s_1s_1} + H_2H_4^*\Phi_{s_2s_2}|^2}{(|H_1|^2\Phi_{s_1s_1} + |H_2|^2\Phi_{s_2s_2}) \cdot (|H_3|^2\Phi_{s_1s_1} + |H_4|^2\Phi_{s_2s_2})} \\
 &= \frac{|1 \cdot 2 \cdot \Phi_{s_1s_1} + 0.1 \cdot 3 \cdot \Phi_{s_2s_2}|^2}{(|1|^2\Phi_{s_1s_1} + |0.1|^2\Phi_{s_2s_2}) \cdot (|2|^2\Phi_{s_1s_1} + |4|^2\Phi_{s_2s_2})} \\
 &= \frac{4 \cdot \Phi_{s_1s_1}^2 + 1.2 \cdot \Phi_{s_1s_1} \Phi_{s_2s_2} + 0.09 \cdot \Phi_{s_2s_2}^2}{4 \cdot \Phi_{s_1s_1}^2 + 9.04 \cdot \Phi_{s_1s_1} \Phi_{s_2s_2} + 0.09 \cdot \Phi_{s_2s_2}^2}
 \end{aligned}$$

und weist damit einen Wert kleiner Eins auf. Angenommen $s_1(t)$ und $s_2(t)$ stellen jeweils die Nutz- bzw. Störsignalquelle dar, müssen die Koeffizienten der mehrkanaligen Signalverarbeitung so gewählt werden, daß bei unverzerrter Übertragung von $s_1(t)$ das Signal $s_2(t)$ vollständig unterdrückt wird. Aufgrund der bekannten Übertragungsfunktionen läßt sich diese Vorgabe mit dem Koeffizientenvektor

$$\begin{aligned}
 \mathbf{w} &= \left[\frac{H_4}{H_1H_4 - H_2H_3} \quad \frac{-H_2}{H_1H_4 - H_2H_3} \right]^T \\
 &= \left[\frac{3}{3 - 0.2} \quad \frac{-0.1}{3 - 0.2} \right]^T
 \end{aligned}$$

realisieren. Damit berechnet sich das Ausgangssignal $y(t)$ zu

$$\begin{aligned}
 y(t) &= \mathbf{w}^T * \mathbf{x} \\
 &= [3/2.8 \quad -0.1/2.8] * [x_1(t) \quad x_2(t)]^T \\
 &= 3/2.8 \cdot (s_1(t) + 0.1 \cdot s_2(t)) - 0.1/2.8 \cdot (2 \cdot s_1(t) + 3 \cdot s_2(t)) \\
 &= s_1(t).
 \end{aligned}$$

An diesem Beispiel wird zum einen das Prinzip der mehrkanaligen Störgeräuschunterdrückung gerichteter Signale deutlich: Die hohe Störsignaldämpfung wird durch das Auslösen einzelner Quellen erreicht. Zum anderen läßt sich feststellen, daß trotz verringerter Kohärenz eine maximale Störsignaldämpfung erreicht werden kann.

Die Ursache für die verringerte Kohärenz bzw. MSC ist, daß diese stets zwischen den Sensorsignalen berechnet wird, also immer nur die gesamte Überlagerung der einzelnen Signalquellen in die Berechnung bzw. in die Messung einfließt. Für die Beurteilung der erreichbaren Störsignaldämpfung ist jedoch vielmehr entscheidend, ob die einzelnen Komponenten des Signalgemisches an den Sensoren kohärent auftreten. Erst dann kann davon ausgegangen werden, daß eine Übertragungsfunktion zwischen den Sensoren existiert, deren Kenntnis die vollständige Auslöschung einer einzelnen Komponente ermöglicht. Für gerichtet einfallende Signale ist diese Bedingung im allgemeinen erfüllt.

2.1.3 Unterdrückung inkohärenter Störsignale

Wie bereits am Anfang von Abschnitt 2.1 erwähnt, ist die erreichbare Störsignaldämpfung für inkohärente bzw. unkorrelierte Signale begrenzt. Das im vorigen Abschnitt beschriebene Prinzip der Auslöschung einzelner Quellen ist auf unkorrelierte Sensorsignale nicht übertragbar, da das jeweilige Störsignal nur an einem einzelnen Sensor anliegt. Es läßt sich also kein Übertragungsfaktor zwischen den Sensoren finden, der für das vollständige Auslösen nötig wäre.

Aus diesem Grund können unkorrelierte Sensorsignale bei mehrkanaliger Signalverarbeitung direkt nur durch das Prinzip der Mittelung gedämpft werden. Dabei wird die Tatsache ausgenutzt, daß sich kohärente Signale bei entsprechendem Phasenausgleich in ihrer Amplitude addieren, während sich bei unkorrelierten Signalen nur die Leistungen addieren. Zur Verdeutlichung sei die mehrkanalige Signalerfassung mit M Mikrofonen betrachtet, deren i -tes Mikrofonsignal als

$$x_i(t) = s(t) + n_i(t) \tag{2.35}$$

dargestellt werden kann, wenn das Nutzsignal $s(t)$ an allen Sensoren mit gleicher Amplitude und Phase anliegt. Die Addition aller Mikrofonsignale, normiert auf die Anzahl der

Mikrofone, ergibt das Ausgangssignal

$$y(t) = \frac{1}{M} \sum_{i=1}^M x_i(t) \quad (2.36)$$

$$= \frac{1}{M} \sum_{i=1}^M (s(t) + n_i(t)) \quad (2.37)$$

$$= s(t) + \frac{1}{M} \sum_{i=1}^M n_i(t). \quad (2.38)$$

Die Leistung des Ausgangssignals berechnet sich unter der Voraussetzung der Unkorreliertheit aller Rauschsignale zu

$$E[y^2] = \sigma_s^2 + \frac{1}{M^2} \sum_{i=1}^M \sigma_{n_i}^2. \quad (2.39)$$

Setzt man für das Rauschen an den Sensoren jeweils die gleiche Leistung σ_n^2 an, vereinfacht sich Gleichung (2.39) wie folgt

$$E[y^2] = \sigma_s^2 + \frac{\sigma_n^2}{M}. \quad (2.40)$$

Folglich wird bei am Ein- und Ausgang gleicher Quellsignalleistung das unkorrelierte Rauschen mit

$$\frac{\sigma_n^2}{\sigma_{n_y}^2} = M \quad (2.41)$$

bzw. mit

$$10 \log_{10} M \text{ [dB]} \quad (2.42)$$

gedämpft. Die sich daraus ergebende Kurve ist in Abbildung 2.5 für bis zu 32 Sensoren dargestellt. Es wird deutlich, daß eine Verdoppelung der Anzahl an Sensoren eine Verbesserung der Störsignaldämpfung um 3 dB zur Folge hat. Für die meisten Anwendungen ist diese Art der Störsignalunterdrückung ungünstig, da die erforderliche Anzahl an Mikrofonen aus Platz- und Kostengründen zu hoch ist.

Trotzdem sind mehrkanalige Verfahren auch bei unkorreliertem Rauschen vorteilhaft einsetzbar, da sie eine zuverlässige Schätzung des Leistungsdichtespektrums des stets korrelierten Nutzsignalanteils erlauben. Die Schätzung beruht auf der Bestimmung der Kreuzkorrelation zwischen den verschiedenen Sensorsignalen. Ausgehend vom Modell nach Gleichung (2.14) und (2.15), bei dem die Mikrofonensignale aus gefiltertem Quellsignal und unkorreliertem Rauschen bestehen, liefert die Kreuzkorrelation im Frequenzbereich

$$E[X_1(\omega)X_2^*(\omega)] = H_1 H_2^* \Phi_{ss}(\omega). \quad (2.43)$$

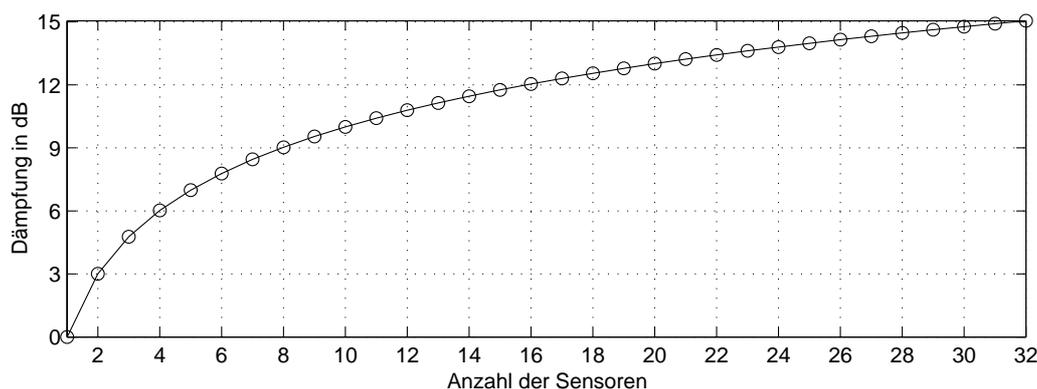


Abbildung 2.5: Störgeräuschunterdrückung bei unkorreliertem Rauschen

Die Schätzung des Nutzspektrums ist die Grundlage für einkanalige adaptive Nachfilter, die sich am Wienerfilter

$$H_{Wiener}(\omega) = \frac{\Phi_{ss}(\omega)}{\Phi_{ss}(\omega) + \Phi_{nn}(\omega)} \quad (2.44)$$

orientieren (siehe z.B. [107]) und lediglich im Detail an spezielle Rauschbedingungen [28] oder Arraystrukturen angepaßt sind [13],[85].

2.2 Praktische Gesichtspunkte

Auch wenn die Prinzipien der mehrkanaligen Störgeräuschunterdrückung klar nach kohärenten und inkohärenten Signalen unterschieden werden können, läßt sich in der Realität nicht immer eine eindeutige Zuordnung finden. Während Sensorrauschen praktisch stets unkorreliert auftritt, sind gerichtete Quellen nicht uneingeschränkt als korreliert zu betrachten. Die Gründe dafür sind in der blockweisen Signalverarbeitung, der großen Anzahl an Rauschquellen oder auch in Nichtlinearitäten im Übertragungspfad [35] zu suchen. Die beiden wichtigsten Gründe seien nachfolgend näher betrachtet.

Einfluß kurzer Segmentlängen

Sehr viele Methoden der digitalen Signalverarbeitung sind in ihrer Funktion auf die Bearbeitung von Signalsegmenten angewiesen. Insbesondere frequenzbereichsbasierte Algorithmen sind auf die Bearbeitung vergleichsweise kurzer Blöcke beschränkt, da die Blocklänge der Anzahl der für die Diskrete Fouriertransformation (DFT) benutzten Abtastwerte entspricht.

Die reale Signalenstehung und -übertragung ist jedoch ein fortlaufender Prozeß, bei dem, wie bereits in Abschnitt 2.1.1 verdeutlicht, sämtliche Signalquellen nur in gefilterter Form erfaßt werden können. Die Filter stellen dabei die jeweilige Raumimpulsantwort zwischen Quelle und Mikrofon dar, deren Dauer von den akustischen Raumbedingungen abhängt und nicht selten im Bereich von mehr als 100 ms liegt. Das bedeutet, daß bei einer Abtastfrequenz von z.B. 22 kHz mindestens 2200 Filtertaps für die diskrete Modellierung eines Filters verwendet werden müssen. Tatsächlich liegt die benutzte Filterordnung aber oft weit unter diesem Wert.

Da sich die meisten Störschallquellen oft außerhalb des Hallradius befinden, wo der Anteil des Nachhalls gegenüber dem Direktschall dominiert, weisen sie sehr lange Impulsantworten auf. Der von der modellierten Filterordnung nicht erfaßte Signalanteil wirkt sich dann korrelationsmindernd aus. Auf diesem Prinzip beruhen beispielsweise adaptive Nachfilter zur Verminderung der Verhallung, bei denen der Nachhallanteil aufgrund der oben beschriebenen Zusammenhänge als unkorreliert angenommen wird [95].

Einfluß der Anzahl der Quellen

In vielen Aufnahmesituationen hat man es mit einer Vielzahl von unabhängigen, gerichteten Störschallquellen zu tun. Diese überlagern sich an den Sensoren, wodurch die Kohärenz, wie in Gleichung (2.23) gezeigt, verringert wird. Im Extremfall von unendlich vielen unabhängigen Rauschquellen entsteht das diffuse Rauschfeld mit si-förmiger Kohärenzfunktion nach Gleichung (2.6) auf Seite 5.

Im Fall von gerichteten Signalen wäre zumindest für eine endliche Anzahl an Quellen die vollständige Auslöschung möglich. Allerdings sind dafür wenigstens so viele Mikrofone wie Quellen nötig, was in der Praxis kaum der Fall sein wird. Hinzu kommt, daß bestimmte Annahmen für die Überlagerung gelten müssen, wie z.B. daß sich die einzelnen Übertragungsfunktionen hinreichend unterscheiden, damit der Prozeß der Überlagerung rückgängig gemacht werden kann. Andernfalls entstehen numerische Probleme in der Stufe zur mehrkanaligen Signalverarbeitung.

Diese Tatsachen führen dazu, daß man in der Praxis oft das stark unterbestimmte Gleichungssystem der Überlagerung der hier für zwei Mikrofonensignale dargestellten Form

$$x_1(t) = h_1 * s_1(t) + h_2 * s_2(t) + \dots + h_N * s_N(t) \quad (2.45)$$

$$x_2(t) = h_{N+1} * s_1(t) + h_{N+2} * s_2(t) + \dots + h_{2N} * s_N(t) \quad (2.46)$$

in ein weniger unterbestimmtes Gleichungssystem umwandelt. Geht man von einem einzigen Nutzsignal aus und faßt den Rest der Quellen plus unkorreliertem Sensorrauschen

zusammen, ergibt sich das aus Gleichung (2.14) und (2.15) bekannte Modell der Mikrofon-signale

$$x_1(t) = \tilde{h}_1 * s(t) + n_1(t) \quad (2.47)$$

$$x_2(t) = \tilde{h}_2 * s(t) + n_2(t). \quad (2.48)$$

Allerdings ist auch dieses Modell nur eine Approximation der tatsächlichen Verhältnisse, da die Rauschanteile $n_1(t)$ und $n_2(t)$ nicht vollständig unkorreliert sind, sondern insbesondere bei tiefen Frequenzen stark korreliert sein können. Des weiteren ist es natürlich möglich, mehr als ein Nutzsignal in die Betrachtungen zu integrieren.

Abschließend sei bemerkt, daß neben klassischen Wienerfiltern als Nachfilter auch andere einkanalige Verfahren, wie z.B. die Spektrale Subtraktion [96] oder Ephraim-Malah-Filter [36], [37] eingesetzt werden können, die z.T. sehr gute Ergebnisse bei der Rauschunterdrückung erzielen.

3 Beamforming-Verfahren

3.1 Theoretische Grundlagen

Um einen Einblick in existierende Beamforming-Verfahren zu bekommen, müssen einige theoretische Grundlagen näher betrachtet werden. Dazu gehört die Ausbreitung der Schallwellen und die Erfassung der Wellen durch Mikrofonarrays mit den entsprechenden Vereinfachungen für Linienarrays. Weiterhin lassen sich die meisten der behandelten Verfahren vorteilhaft im Frequenzbereich realisieren, weshalb im nächsten Abschnitt zunächst die diskrete Kurzzeit-Fouriertransformation thematisiert wird.

3.1.1 Blockweise Verarbeitung

Eine weit verbreitete Möglichkeit, die Anzahl der gleichzeitig zu adaptierenden Parameter zu verkleinern, stellt die Verarbeitung in Teilbändern dar. Wählt man für die Realisierung der Bänder eine diskrete Kurzzeit-Fouriertransformation, lassen sich die Faltungsoperationen im Zeitbereich als komplexwertige Multiplikationen im Frequenzbereich realisieren. Voraussetzung dafür ist, daß die an sich bei Multiplikationen im Frequenzbereich zirkulare Faltung als lineare Faltung dargestellt werden kann.

Eine Annäherung an diese Forderung wird in Anlehnung an die *Overlap-Save*-Methode durch überlappende Blockbildung und Gewichtung mit einer geeigneten Fensterfunktion erreicht. Die Fensterung erfolgt in diesem Fall jedoch sowohl zur Analyse als auch zur Synthese, da nur so Randeffekte optimal unterdrückt werden können.

Bei einem Versatz V zwischen den fortlaufenden Fenstern und einer Block- bzw. Fensterlänge L weist das nachfolgend definierte Rechteckfenster

$$w_r(t) = \begin{cases} \sqrt{\frac{V}{L}} & 0 \leq t < L \\ 0 & \text{sonst} \end{cases} \quad (3.1)$$

die Eigenschaft auf, sich in der quadrierten Summe zu Eins zu addieren [41]. Das heißt, es gilt

$$\sum_{m=-\infty}^{\infty} w_r^2(mV - t) = \sum_{m=0}^{(L/V)-1} \frac{V}{L} = 1 \quad (3.2)$$

für alle Zeitpunkte t , wobei m den fortlaufenden Index bzw. die Nummer des Fensters bezeichnet. Damit ist sichergestellt, daß ein Signal unverzerrt hin- und wieder rücktransformiert werden kann.

Verändert man jedoch das transformierte Signal im Frequenzbereich durch Multiplikation mit Filterkoeffizienten, können aufgrund der äquivalenten Faltungsoperation im Zeitbereich Randeffekte auftreten. Diese lassen sich mit Hilfe des Rechteckfensters $w_r(t)$ nicht unterdrücken. Aus diesem Grund wird eine sinusförmige Fensterfunktion mit einer zum Rand abfallenden Gewichtung genutzt.

Für $L = k \cdot 4V$, also wenn die Blocklänge ein Vielfaches des vierfachen Fensterversatzes ist, erfüllt das Fenster

$$w_H(t) = \frac{2w_r(t)}{\sqrt{4a^2 + 2b^2}} \left[a + b \cos\left(\frac{2\pi t}{L} + \Phi\right) \right] \quad (3.3)$$

ebenfalls die Anforderung aus Gleichung (3.2). Bei der Wahl von $a = 0.5$, $b = -0.5$ und $\Phi = \pi/L$ entsteht ein modifiziertes Hanning-Fenster, das in jedem Block zur Berechnung des gefensterten Signals $x_w(t)$ verwendet wird [41]

$$x_w(mV, t) = w_H(mV - t)x(t). \quad (3.4)$$

Mit Hilfe der diskreten Fouriertransformation lassen sich die einzelnen Blöcke der Länge L in den Frequenzbereich überführen

$$x_w(mV, t) \quad \circ \bullet \quad X(mV, \omega), \quad (3.5)$$

so daß die Kurzzeit-Fouriertransformierte (*Short Time Fourier Transform (STFT)*) entsteht. Die explizite Abhängigkeit von der Zeit bzw. der Blocknummer wird häufig nicht weiter mitgeführt, so daß für alle Verfahren, die auf der STFT beruhen, die Notation $X(\omega) = X(\omega, mV) = X(\omega, \tilde{t})$ mit $\tilde{t} = mV$ gelten soll.

Die Rekonstruktion der mit Hilfe der inversen Fouriertransformation aus dem Frequenz- in den Zeitbereich rücktransformierten Signalblöcke $y_w(mV, t)$ zu einem fortlaufenden Signal erfolgt schließlich durch Addition der jeweils mit w_H gewichteten Blöcke

$$y(t) = \sum_{m=-\infty}^{\infty} w_H(mV - t) \cdot y_w(mV, t). \quad (3.6)$$

3.1.2 Ausbreitung von Schallwellen

In der Physik wird die räumliche Ausbreitung von Wellen in Abhängigkeit vom jeweiligen Medium durch Wellengleichungen mit Randbedingungen beschrieben. Für ein skalares Wellenfeld eines Signals $s(x, y, z, t)$ am Ort (x, y, z) zum Zeitpunkt t gilt die Wellengleichung [49]

$$\frac{\partial^2 s}{\partial x^2} + \frac{\partial^2 s}{\partial y^2} + \frac{\partial^2 s}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 s}{\partial t^2}. \quad (3.7)$$

Dabei stellt c die mediumspezifische Ausbreitungsgeschwindigkeit dar, die in der Luft unter anderem vom Druck und von der Temperatur abhängig ist. Unter der Voraussetzung eines homogenen Raumes und von Schallwellen im hörbaren Bereich kann bei Normaldruck und einer Lufttemperatur von 20°C für c ein konstanter Wert von $343 \frac{\text{m}}{\text{s}}$ angenommen werden [89].

Gleichung (3.7) stellt eine partielle Differentialgleichung zweiter Ordnung dar. Geht man von der Existenz separierbarer Lösungen der Form

$$s(x, y, z, t) = f(x) \cdot f(y) \cdot f(z) \cdot f(t)$$

aus, kann das Signal $s(x, y, z, t)$ in kartesischen Koordinaten angesetzt werden als

$$s(x, y, z, t) = A \exp\{i(\omega t - k_x x - k_y y - k_z z)\}, \quad (3.8)$$

mit A als komplexwertiger Konstante, $\{k_x, k_y, k_z\} \in \mathbb{R}$, $\omega = 2\pi f = \text{const} \in \mathbb{R}$ und $\omega \geq 0$. In (3.7) eingesetzt ergibt sich

$$k_x^2 s(x, y, z, t) + k_y^2 s(x, y, z, t) + k_z^2 s(x, y, z, t) = \frac{\omega^2 s(x, y, z, t)}{c^2}, \quad (3.9)$$

woraus für die Parameter in Gleichung (3.8) die Beschränkungen

$$k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c^2} \quad (3.10)$$

folgen. Mit der Einführung des Wellenzahlvektors $\mathbf{k} = [k_x \ k_y \ k_z]$ und des Positionsvektors $\mathbf{p} = [x \ y \ z]^T$ läßt sich Gleichung (3.8) in kompakter Form als

$$s(x, y, z, t) = A \exp\{i(\omega t - \mathbf{k}\mathbf{p})\} \quad (3.11)$$

darstellen. Diese Gleichung beschreibt eine monochromatische planare Welle der Frequenz ω , die sich ungedämpft mit einer ebenen Wellenfront in Richtung \mathbf{k} ausbreitet.

Häufig gebraucht man zur Beschreibung des Wellenfeldes zusätzlich auch die Wellenlänge, also die Strecke, die eine Ebene konstanter Phasenlage in der Zeit $2\pi/\omega$ zurücklegt. In Abhängigkeit von der Ausbreitungsgeschwindigkeit ergibt sie sich zu

$$\lambda = \frac{2\pi c}{\omega} = \frac{c}{f}. \quad (3.12)$$

Mit ihrer Hilfe kann der Betrag des Wellenzahlvektors \mathbf{k} mit der Vorgabe aus Gleichung (3.10) als

$$|\mathbf{k}| = \frac{2\pi}{\lambda} \quad (3.13)$$

angegeben werden. Der Betrag des Wellenzahlvektors gibt damit die Anzahl der Wellenzyklen in Ausbreitungsrichtung mit der Einheit Radiant pro Meter an [49].

Für sich ausbreitende Schallwellen wird in dieser Arbeit stets die Beschreibung der planaren, monochromatischen Welle nach Gleichung (3.11) benutzt. Sie korrespondiert mit der Vorstellung vom Fernfeld-Modell mit planaren Wellenfronten und ermöglicht die Beschreibung beliebiger Signale, da die Wellengleichung linear ist, durch komplexe Exponentialfunktionen erfüllt wird und da beliebige Funktionen als gewichtete Überlagerung komplexer Exponentialfunktionen dargestellt werden können [49].

Abschließend sei angemerkt, daß man die Wellengleichung u.a. auch in Kugelkoordinaten formulieren kann. Die Lösung dieser Gleichung führt dann zur Beschreibung sphärischer, monochromatischer Wellen, wie sie unter Nahfeldbedingungen vorzufinden sind. In Abhängigkeit vom Radius r ergibt sich eine gedämpfte Exponentialfunktion

$$s(r, t) = \frac{A}{r} \exp\{i\omega(t - r/c)\}. \quad (3.14)$$

In Kapitel 5 wird jedoch gezeigt, warum die explizite Unterscheidung zwischen Nah- und Fernfeld im Kontext dieser Arbeit vernachlässigbar ist.

3.1.3 Mikrofonarrays

Mikrofonarrays dienen der räumlichen Erfassung von Schallfeldern durch Abtastung an diskreten Stellen im Raum. Auch wenn es eine Vielzahl vorstellbarer Arraygeometrien gibt (siehe z.B. [93]), beschränken sich die Ausführungen in dieser Arbeit auf äquidistante Linienarrays. Dies ist unter anderem darin begründet, daß bei dem in Kapitel 5 vorgestellten Algorithmus zwei Mikrofone zugrunde gelegt sind, die als Sonderfall eines Linienarrays angesehen werden können. Abbildung 3.1 zeigt ein solches Linienarray mit einfallender planarer Schallwelle.

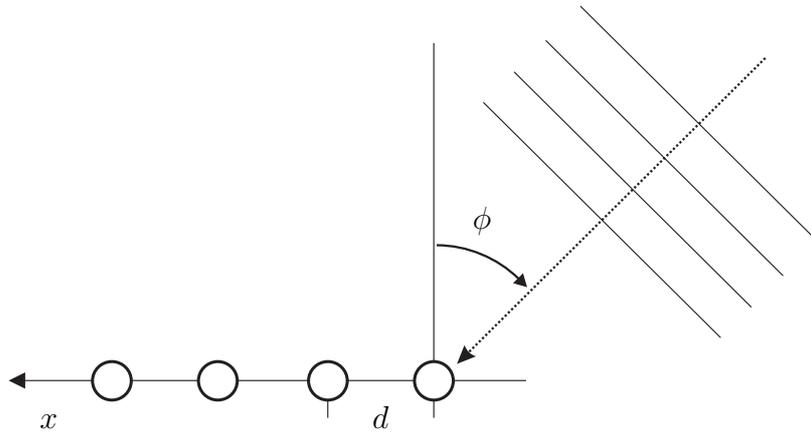


Abbildung 3.1: Äquidistantes Linienarray zur mehrkanaligen Signalerfassung

Da es sich um ein eindimensionales Array entlang der x-Achse handelt, kann die Beschreibung der Wellenausbreitung auf die Komponente in x-Richtung reduziert werden. Diese ergibt sich zu

$$k_x = |\mathbf{k}| \sin(\phi) = \frac{2\pi}{\lambda} \sin(\phi). \quad (3.15)$$

Weiterhin kann die Erfassung einer sich ausbreitenden Welle mit Hilfe eines Arrays in Bezug auf einen Referenzsensor formuliert werden, da man zur Störsignalunterdrückung nur auf die relativen Signalunterschiede zwischen den einzelnen Sensoren angewiesen ist (vgl. Abschnitt 4.3, Seite 64).

Für das im Koordinatenursprung liegende Referenzmikrofon kann damit also eine Amplitude von $A = 1$ zum Zeitpunkt $t = 0$ angenommen werden, woraus für die Welle folgt

$$s(x = 0, t = 0) = \exp\{i\omega \cdot 0\} = 1. \quad (3.16)$$

Im Vergleich dazu befindet sich das zweite Mikrofon in einem Abstand d in x-Richtung versetzt. Die Welle muß also an der Stelle $x = d$ zum Zeitpunkt $t = 0$ betrachtet werden, indem die Ausbreitung in x-Richtung aus Gleichung (3.15) eingesetzt wird

$$\begin{aligned} s(x = d, t = 0) &= \exp\{i(\omega \cdot 0 - k_x \cdot d)\} \\ &= \exp\left\{-i2\pi \frac{d}{\lambda} \sin(\phi)\right\} \\ &= \exp\left\{-i\omega \frac{d}{c} \sin(\phi)\right\}. \end{aligned} \quad (3.17)$$

Bezogen auf das Referenzmikrofon weist die sich frei ausbreitende, ungedämpfte, ebene Welle an der Stelle des zweiten Sensors einen Phasenversatz nach Gleichung (3.17) auf.

Beim Einsatz omnidirektionaler Mikrofone lassen sich diese Phasenfaktoren zur Modellbildung nutzen (siehe Abschnitt 5.2.2 ab Seite 73). Für M Mikrofone ergeben sich die M Phasenfaktoren zu

$$r_k(\omega, \phi) = \exp\left\{-i\omega \frac{d_k}{c} \sin(\phi)\right\} \quad \text{mit } k = 1 \dots M \text{ und } d_k = (k - 1) \cdot d. \quad (3.18)$$

Sie lassen sich im Vektor

$$\mathbf{r}(\omega, \phi) = [1, r_2(\omega, \phi), \dots, r_M(\omega, \phi)]^T \quad (3.19)$$

als Array-Antwort auf räumliche Wellen zusammenfassen. Der Vektor $\mathbf{r}(\omega, \phi)$ wird auch als Steering-Vektor bezeichnet und z.B. zur Berechnung des Beampatterns (siehe Abschnitt 3.2.2.1) oder bei adaptiven Beamformern für die Formulierung von Nebenbedingungen verwendet (siehe Abschnitt 3.4).

Wie zu Beginn dieses Abschnitts betont, läßt sich die fortlaufende Welle, bedingt durch die diskreten Mikrofonpositionen, nicht kontinuierlich erfassen sondern wird abgetastet. Dieser Prozeß weist viele Ähnlichkeiten zur Abtastung im Zeitbereich auf. So lassen sich, analog zur Fensterung von Zeitsignalen, unterschiedliche Sensorgewichtungen anwenden, deren spektrale Eigenschaften, wie Breite der Hauptkeule und Dämpfung der Nebenkeulen, sich in diesem Fall auf die mehrdimensionale räumliche Fouriertransformierte auswirken.

Von besonderer Bedeutung ist im Kontext dieser Arbeit das Problem der Unterabtastung. Auch das räumliche Spektrum einer abgetasteten Welle weist periodische Fortsetzungen bei Vielfachen von $2\pi/d$ auf, wobei der Sensorabstand d die räumliche Abtastfrequenz darstellt. Es kann gezeigt werden [93], daß diese Fortsetzungen im sichtbaren Bereich des Spektrums auftreten können, wenn nicht die Bedingung

$$d \leq \frac{\lambda_{f_{max}}}{2} \quad (3.20)$$

eingehalten wird. Dabei stellt $\lambda_{f_{max}}$ die Wellenlänge der höchsten im Signal vorkommenden Frequenz dar. Mit Hilfe der Beziehung $c = \lambda \cdot f$ kann (3.20) umformuliert werden zu

$$d \leq \frac{c}{2f_{max}}. \quad (3.21)$$

Die bei Verletzung dieser Bedingung in den sichtbaren Bereich wandernden Wiederholungen der Hauptkeulen des periodischen Spektrums werden als *Grating Lobes* bezeichnet. Sie haben die gleiche Amplitude wie die ursprüngliche Hauptkeule und bewirken, daß sich Signale aus dieser Einfallsrichtung nicht von Signalen aus der Hauptrichtung unterscheiden lassen.

Für die Verarbeitung breitbandiger Signale mit Beamformern müssen die Sensorabstände also entsprechend Gleichung (3.21) angepaßt sein. Sollen z.B. Sprachsignale mit einer Abtastrate von 11 kHz verarbeitet werden, dürfte der Sensorabstand d einen Wert von 3.09 cm nicht überschreiten.

Abschließend sei zur Sensorplatzierung noch angemerkt, daß die Beschränkung des Sensorabstandes in der o.g. Art und Weise zu einer räumlichen Überabtastung aller Frequenzen $f < f_{max}$ führt, so daß sich Wellen verschiedener Einfallsrichtungen insbesondere in den unteren Frequenzbändern kaum unterscheiden (siehe Abschnitt 5.2.5.3). Aus diesem Grund existieren Ansätze für Breitbandbeamformer, die unterschiedliche Sensorabstände für die verschiedenen Frequenzbereiche vorschlagen [101, 90].

3.2 Einteilung und Bewertung von Beamformern

Beamformer für Mikrofon-signale können sich sowohl in ihrer Struktur als auch in der Art und Weise der Parameteradaption unterscheiden. Um sie miteinander vergleichen zu können, werden in den nachfolgenden beiden Abschnitten die unterschiedlichen Merkmale herausgestellt und Kenngrößen eingeführt, die zur Bewertung der Leistungsfähigkeit nützlich sind.

3.2.1 Einteilung von Beamformern

Ein klassische Möglichkeit der Einteilung von Beamformern ist die Unterscheidung zwischen adaptiven und nichtadaptiven Verfahren, teilweise auch als statistisch optimal bzw. datenunabhängig bezeichnet [97].

Nichtadaptive Beamformer besitzen feste, voreingestellte Parameter, d.h. Nutz- und Stör-signalrichtungen müssen im Vorfeld bekannt sein. Sie weisen eine *Feed-Forward*-Struktur auf und können als Summation von zeitverzögerten Sensorsignalen (siehe Abschnitt 3.3.1 zum Delay-and-Sum-Beamformer) oder als Summation von gefilterten Sensorsignalen (siehe Abschnitt 3.3.2 zum Filter-and-Sum-Beamformer) implementiert werden. Letzteres ermöglicht eine bessere Unterdrückung breitbandiger Störschallquellen, wie z.B. Störsprechen.

Adaptive Beamformer sind in der Lage, sich durch Nachregeln der Parameter an die jeweiligen akustischen Bedingungen anzupassen. Klassische Beamforming-Verfahren gehen

dabei von einer einzelnen Nutzsinalquelle aus, deren Position oft als bekannt vorausgesetzt wird. Um die Auslöschung des Nutzsignals zu vermeiden, werden entweder Nebenbedingungen formuliert (siehe Abschnitt 3.4.1 zum Linear Constrained Minimum Variance Beamformer) oder nutzsinalfreie Störreferenzen erzeugt (siehe Abschnitt 3.4.2 zum Generalized Sidelobe Canceller).

3.2.2 Bewertung von Beamformern

Für die Bewertung von Beamformern haben sich verschiedene Kenngrößen, wie *Beam pattern*, *Array Gain*, *White Noise Gain*, *Directivity* und *Front To Back Ratio* etabliert [14]. Im Rahmen dieser Arbeit sind jedoch lediglich *Beam pattern*, *Array Gain* und *White Noise Gain* von Bedeutung und werden nachfolgend näher betrachtet.

3.2.2.1 Beampattern

Als *Beampattern*, Ortsfrequenzgang oder auch Richtdiagramm wird die graphische Darstellung der frequenz- und ortsabhängigen Dämpfung eines Beamformers bezeichnet. Da die hier betrachteten Linienarrays ein rotationssymmetrisches Dämpfungsverhalten aufweisen, beschränkt sich die Ortsabhängigkeit auf die Abhängigkeit von der Einfallsrichtung ϕ .

Um das Beampattern zu berechnen, wird in jedem Frequenzband die Antwort des Beamformers auf eine ebene Welle, definiert über den relativen Phasenversatz nach Gleichung (3.18), ausgewertet. Die Multiplikation mit den Beamformerkoeffizienten $\mathbf{w}(\omega)$ und die anschließende Summation kann als Skalarprodukt der Form

$$\Psi(\omega, \phi) = \mathbf{w}^H(\omega) \mathbf{r}(\omega, \phi) \quad (3.22)$$

geschrieben werden. Exemplarisch zeigt Abbildung 3.2 auf der nächsten Seite den Betrag eines auf diese Weise berechneten Beampatterns für einen Beamformer mit vier gleichgewichteten Koeffizienten, d.h. mit

$$w_k = \frac{1}{M} = \frac{1}{4} \quad \text{und} \quad k = 1 \dots M. \quad (3.23)$$

Üblicherweise erfolgt die Darstellung der frequenz- und blickrichtungsabhängigen Dämpfung in logarithmierter Form unter Angabe der entsprechenden dB-Werte. Neben der zweidimensionalen Darstellung aus Abbildung 3.2 ist auch die dreidimensionale Darstellung wie in Abbildung 3.3 auf Seite 24 sowie die Ansicht als Überlagerung aller Frequenzbänder wie in Abbildung 3.4 auf Seite 25 gebräuchlich.

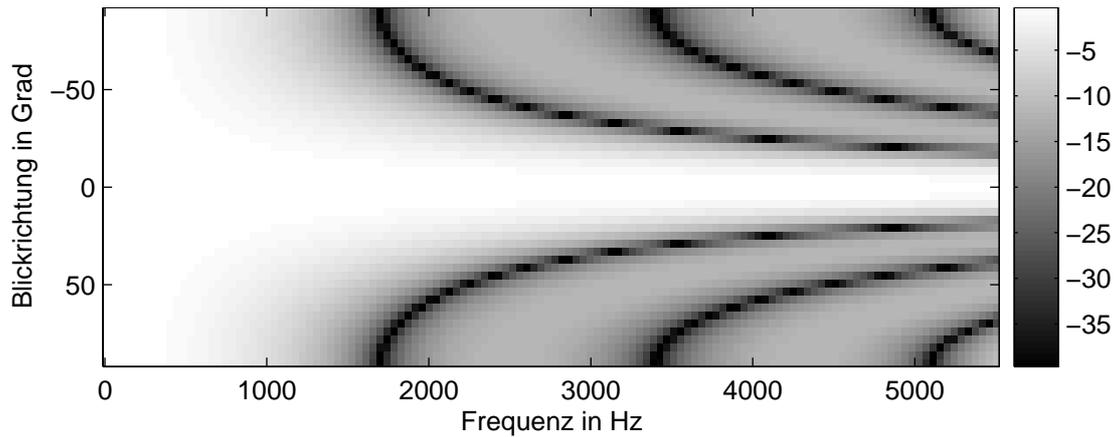


Abbildung 3.2: Beampattern bei vier Koeffizienten mit einer konstanten Gewichtung von $\frac{1}{4}$ und einem Mikrofonabstand von $d = 5$ cm

3.2.2.2 Array Gain

Der Array-Gewinn G (engl. *Array Gain*) gibt den Gewinn bei Verwendung eines Arrays, also die Verbesserung des Signal-Rauschabstandes (engl. *Signal to Noise Ratio*, SNR), im Vergleich zu einem einzelnen Sensor an. Folglich gilt

$$G = \frac{\text{SNR}_{\text{Array}}}{\text{SNR}_{\text{Sensor}}}. \quad (3.24)$$

Das frequenzabhängige SNR an einem einzelnen Sensor ist durch das Verhältnis von Nutzsignalleistung zur mittleren Rauschleistung bestimmt [14]

$$\text{SNR}_{\text{Sensor}} = \frac{\Phi_{ss}}{\Phi_{\bar{n}\bar{n}}}. \quad (3.25)$$

Für die Berechnung des Array-SNR wird die frequenzabhängige Ausgangsleistung des Arrays bei ausschließlicher Nutzsignalanwesenheit zur Ausgangsleistung bei Rauschanwesenheit ins Verhältnis gesetzt

$$\text{SNR}_{\text{Array}} = \frac{\Phi_{ss} \cdot |\mathbf{w}^H \mathbf{r}|^2}{\Phi_{\bar{n}\bar{n}} \mathbf{w}^H \Phi_{\mathbf{nn}} \mathbf{w}}, \quad (3.26)$$

wobei $\Phi_{\mathbf{nn}}$ die normierte Matrix der Sensor-Kreuzleistungsdichtespektren des Rauschens darstellt. Mit Hilfe der Gleichungen (3.25) und (3.26) lässt sich der Array Gewinn als

$$G = \frac{|\mathbf{w}^H \mathbf{r}|^2}{\mathbf{w}^H \Phi_{\mathbf{nn}} \mathbf{w}} \quad (3.27)$$

angeben [14]. Der Zähler in (3.27) wird, da er das Verhalten des Beamformers bei präsenter Nutzschallquelle widerspiegelt, als Signalantwort und der Nenner als Rauschantwort bezeichnet.

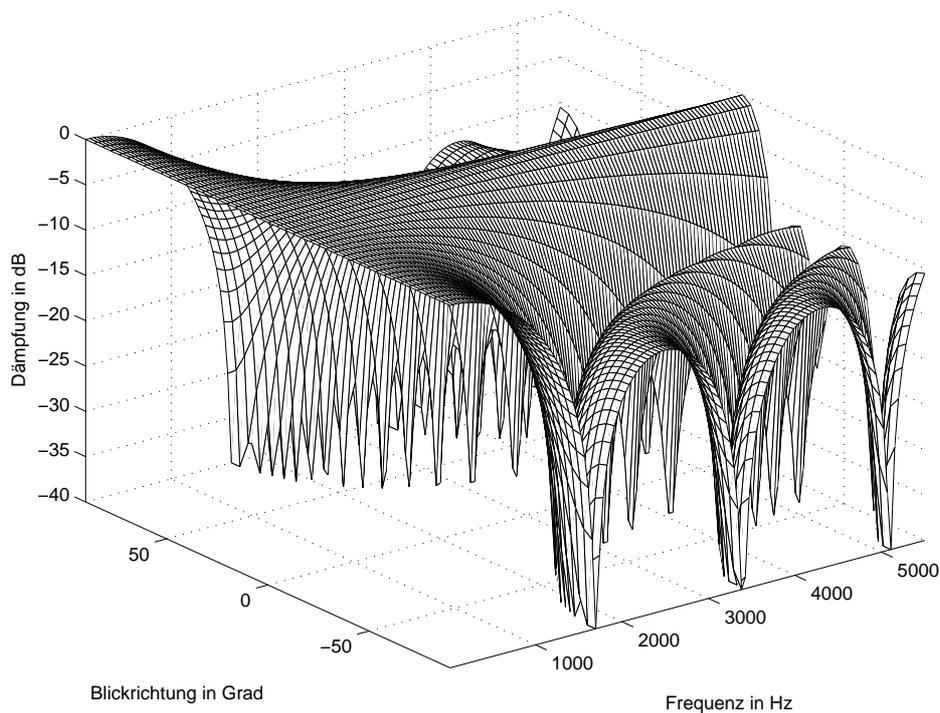


Abbildung 3.3: Beampattern bei vier Koeffizienten mit einer konstanten Gewichtung von $\frac{1}{4}$ und einem Mikrofonabstand von $d = 5$ cm in dreidimensionaler Darstellung

3.2.2.3 White Noise Gain (WNG)

Diese Kenngröße beschreibt, wie sich der Beamformer bei räumlich unkorreliertem Rauschen, wie es z.B. durch das Eigenrauschen der Sensoren gegeben ist, verhält. Das *White Noise Gain* kann auch als Maß für die Robustheit eines Beamformers verstanden werden, da Fehler in der Arrayausrichtung, z.B. verursacht durch falsche Mikrofonpositionen, ebenfalls als unkorrelierte Größen modelliert werden können [29],[40].

Die klassische Definition des White Noise Gain ist aus dem Array-Gewinn abgeleitet. Für unkorreliertes Rauschen ist die normierte Matrix der Kreuzleistungsdichtespektren über die Sensorsignale gegeben als

$$\Phi_{\mathbf{nn}} = \mathbf{I}, \quad (3.28)$$

wobei \mathbf{I} die Einheitsmatrix darstellt. Damit berechnet sich der Array Gewinn entsprechend Gleichung (3.27) zu

$$G_w = \frac{|\mathbf{w}^H \mathbf{r}|^2}{\mathbf{w}^H \mathbf{w}}. \quad (3.29)$$

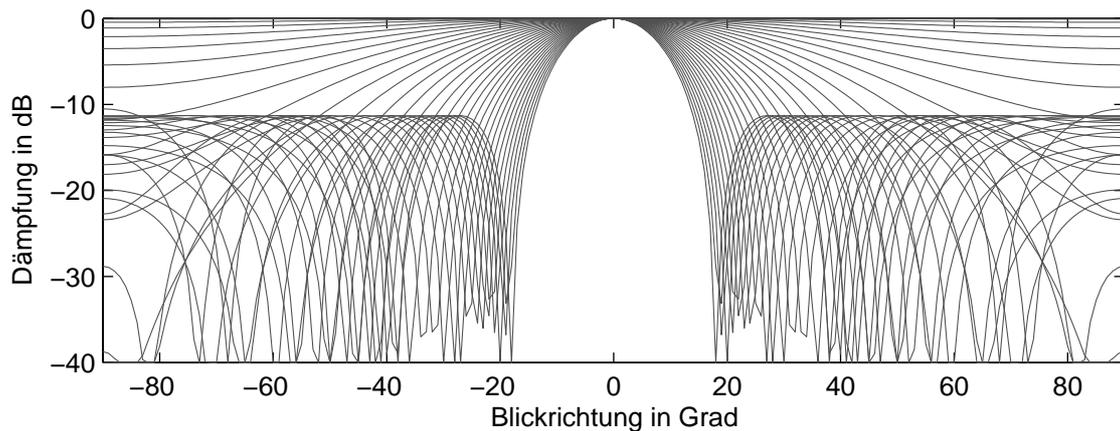


Abbildung 3.4: Beam pattern bei vier Koeffizienten mit einer konstanten Gewichtung von $\frac{1}{4}$ und einem Mikrofonabstand von $d = 5$ cm als Überlagerung aller Frequenzbänder dargestellt

G_w wird als White Noise Gain bezeichnet, weil es das Verhalten des Beamformers bei unkorreliertem Rauschen in Relation zur Nutzsignaldämpfung beschreibt.

Wie in späteren Abschnitten noch gezeigt wird (siehe Abschnitt 5.2.1), ist das Konzept der ungedämpften Nutzsignalübertragung als Funktion von \mathbf{r} in realen Umgebungen fehlerbehaftet. Aus diesem Grund wird nachfolgend eine von der klassischen Definition des WNG abweichende Bewertung der Rauschverstärkung eingeführt, die ausschließlich auf der Norm der Koeffizienten beruht. Zu diesem Zweck soll exemplarisch das Verhalten eines Beamformers mit zwei Mikrofonen im Fall von unkorreliertem Rauschen näher untersucht werden. Grundlage dafür ist das in Abbildung 3.5 dargestellte Modell eines zweikanaligen Beamformers.

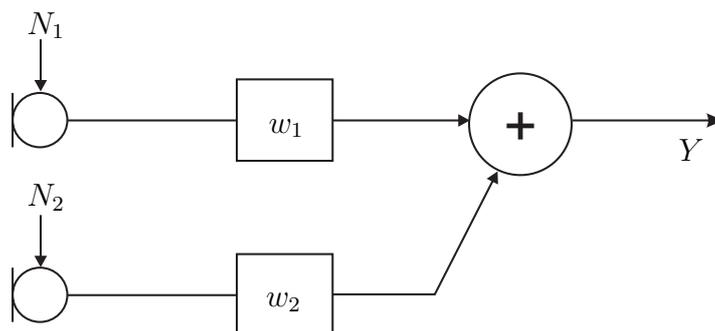


Abbildung 3.5: Zweikanaliger Beamformer mit additivem Sensorrauschen

Die Ausgangsleistung in jedem Frequenzband berechnet sich ohne explizite Berücksichti-

gung der Frequenzabhängigkeit zu

$$\begin{aligned}
 E[|Y|^2] &= E[YY^*] \\
 &= E[(N_1w_1 + N_2w_2)(N_1^*w_1^* + N_2^*w_2^*)] \\
 &= E[w_1w_1^*N_1N_1^*] + E[w_1w_2^*N_1N_2^*] + E[w_2w_1^*N_2N_1^*] + E[w_2w_2^*N_2N_2^*] \\
 &= |w_1|^2E[N_1N_1^*] + |w_2|^2E[N_2N_2^*] \\
 &= |w_1|^2\sigma_{N_1}^2 + |w_2|^2\sigma_{N_2}^2.
 \end{aligned}$$

Setzt man an jedem Mikrofon die gleiche Rauschleistung σ_N^2 an, wird deutlich, daß die Ausgangsleistung des Beamformers die mit der Summe der Betragsquadrate der Koeffizienten verstärkte Rauschleistung ist

$$E[|Y(\omega)|^2] = \sigma_N^2 \cdot (|w_1|^2 + |w_2|^2). \quad (3.30)$$

Im allgemeinen Fall von M Mikrofonen bzw. Koeffizienten pro Frequenzband ergibt sich die Ausgangsleistung zu

$$\sigma_Y^2 = \sigma_N^2 \cdot \mathbf{w}^H \mathbf{w}, \quad (3.31)$$

wobei $\mathbf{w} = [w_1, w_2, \dots, w_M]^T$ den Koeffizientenvektor darstellt. Folglich ist eine geeignete Größe zur Bewertung der Verstärkung unkorrelierten Rauschens durch die euklidische Norm bzw. die quadrierte euklidische Norm

$$\|\mathbf{w}\|_2^2 = \mathbf{w}^H \mathbf{w} \quad (3.32)$$

gegeben. Diese Größe wird im weiteren Verlauf der Arbeit zur Bewertung der Rauschverstärkung herangezogen. Die Darstellung erfolgt in logarithmierter Form unter Angabe der entsprechenden dB-Werte.

3.3 Nichtadaptive Beamformer

Nichtadaptive Beamformer mit fest entworfenen Koeffizienten werden üblicherweise in Strukturen mit einfachen Verzögerungsgliedern und Strukturen mit frequenzabhängigen Filtern unterteilt. Allerdings ist diese Art der Klassifizierung insofern willkürlich, als daß reine Verzögerungsglieder nur eine Untergruppe der allgemeineren Filterstrukturen sind. Beide werden ausschließlich anhand von Phaseninformationen, meist aus dem Fernfeld-Modell stammend, entworfen. Während Beamformer mit Verzögerungsgliedern durch die Vorgabe eines Koeffizientensatzes bei einer Frequenz für alle anderen Frequenzen festgelegt sind, müssen bei Beamformern mit Filterstrukturen alle Frequenzbänder separat entworfen werden.

3.3.1 Delay-and-Sum-Beamformer

Der *Delay-&Sum-Beamformer* nach Abbildung 3.6 ist sowohl der einfachste als auch der robusteste Beamformer. Die räumlichen Informationen bezüglich des Quellsignals sind durch die Laufzeitunterschiede modelliert, so daß sich das Ausgangssignal in allgemeiner Form als

$$y_{DS}(t) = a_1 \cdot x_1(t - \tau_1) + a_2 \cdot x_2(t - \tau_2) + \dots + a_M \cdot x_M(t - \tau_M) \quad (3.33)$$

$$= \sum_{k=1}^M a_k \cdot x_k(t - \tau_k) \quad (3.34)$$

schreiben läßt. Ohne Berücksichtigung der Gesamtverzögerung zwischen Quelle und Array kann Gleichung (3.33) umformuliert werden zu

$$y_{DS}(t) = a_1 \cdot x_1(t) + a_2 \cdot x_2(t - \tilde{\tau}_2) + \dots + a_M \cdot x_M(t - \tilde{\tau}_M). \quad (3.35)$$

Die einzelnen $\tilde{\tau}_k$ beschreiben damit nicht mehr die Laufzeit des Signals zwischen Quelle und jeweiligem Sensor, sondern statt dessen die auf das Referenzmikrofon bezogenen Laufzeitunterschiede.

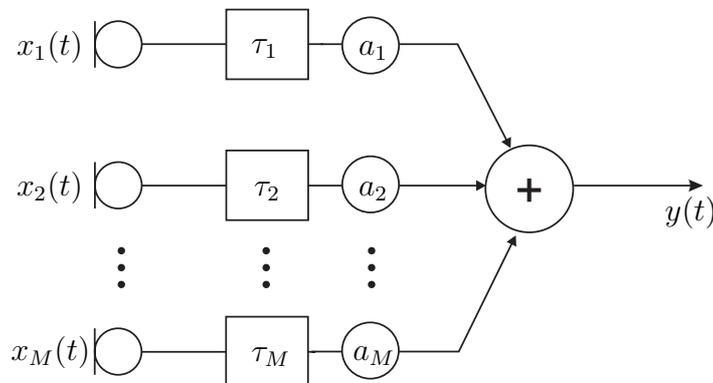


Abbildung 3.6: Allgemeine Struktur des Delay-and-Sum-Beamformers

Häufig sind die einzelnen Koeffizienten a_k gleich gewichtet, womit die räumliche Fensterfunktion auf ein Rechteckfenster festgelegt ist. Um eine Gesamtverstärkung von Eins zu erhalten, müssen die Koeffizienten auf die Sensoranzahl normiert werden. Für das Ausgangssignal folgt damit:

$$y_{DS}(t) = \frac{1}{M} \cdot [x_1(t) + x_2(t - \tilde{\tau}_2) + \dots + x_M(t - \tilde{\tau}_M)]. \quad (3.36)$$

Nachteilig an der Implementierung des Delay-and-Sum-Beamformers im Zeitbereich ist, daß die Verzögerungsglieder in digitalen Systemen stets ganzzahlige Vielfache der Abtast-rate sein müssen. Das bedeutet gleichzeitig, daß die möglichen Einfallrichtungen ebenfalls auf diskrete Werte beschränkt sind. Aus diesem Grund ist eine Implementierung im Frequenzbereich vorteilhaft, wobei die einzelnen Verzögerungsglieder dann jeweils als lineare Phasenverschiebung auftreten.

Entsprechend den Vorgaben aus dem Fernfeld-Modell hängt die an jedem Sensor anzuwendende Phasenverschiebung lediglich von der Einfallrichtung ϕ der Nutzsignalquelle ab. Damit können die Koeffizienten des Delay-and-Sum-Beamformers bei räumlichem Rechteckfenster und Normierung auf die Sensoranzahl wie folgt notiert werden:

$$\mathbf{w}_{DS}(\omega) = \frac{1}{M} \mathbf{r}(\phi, \omega). \quad (3.37)$$

Das Ausgangssignal im Frequenzbereich ergibt sich analog zum Zeitbereich als Summation der mit den Koeffizienten gewichteten Sensorsignale. Formuliert man die Summation wie üblich als Skalarprodukt, lautet die Frequenzbereichsformulierung von Gleichung (3.36)

$$Y(\omega) = \mathbf{w}_{DS}^H(\omega) \mathbf{X}(\omega). \quad (3.38)$$

Folglich ist für die Übertragung der Nutzsignalquelle aus Einfallrichtung ϕ mit

$$Y(\omega) = \mathbf{w}_{DS}^H(\omega) \mathbf{X}(\omega) = \frac{\mathbf{r}^H(\phi, \omega) \mathbf{r}(\phi, \omega)}{M} = 1 \quad (3.39)$$

die ungedämpfte Übertragung gewährleistet. Weiterhin läßt sich feststellen, daß die Norm der Koeffizienten mit

$$\mathbf{w}_{DS}^H \mathbf{w}_{DS} = \frac{\mathbf{r}^H(\phi, \omega) \mathbf{r}(\phi, \omega)}{M^2} = \frac{1}{M} \quad (3.40)$$

für $M > 1$ stets Werte kleiner Eins annimmt. Tatsächlich ist das WNG des Delay-and-Sum-Beamformers in dieser Konfiguration im Vergleich zu allen anderen Beamforming-Verfahren optimal, da entsprechend Gleichung (2.41) die maximale Rauschdämpfung von $10 \log_{10} M$ [dB] erreicht wird (vgl. Abschnitt 2.1.3). Jede andere Konfiguration der Koeffizienten verschlechtert das Verhalten bei unkorreliertem Rauschen.

Das für unkorreliertes Rauschen optimale Dämpfungsverhalten wird allerdings durch den nahezu Totalverlust der Dämpfung kohärenter Störer erkauft, wie sich an Abbildung 3.2 bzw. 3.3 erkennen läßt. Insbesondere im Bereich niedriger Frequenzen, mit Wellenlängen, die sehr viel größer als der Sensorabstand sind, und bei denen sich die Signale an den Sensoren kaum unterscheiden, ist keine Richtwirkung erzielbar. Zusätzlich zu dieser Problematik der überabgetasteten Wellen tritt bei Frequenzen mit $\lambda/2 < d$ räumliches Aliasing auf.

Eine Möglichkeit, diese Probleme zu umgehen, besteht in der Implementierung mit Teilarrays für verschiedene Frequenzbereiche. Nimmt man eine ungefähre Arbeitsbandbreite von $0,25 < d/\lambda \leq 0,5$ für den Delay-and-Sum-Beamformer sowie den für Telefonanwendungen nötigen Frequenzbereich von 300 Hz bis 3.4 kHz an, ergibt sich bei fünf Mikrofonen je Teilband eine Anzahl von 18 notwendigen Mikrofonen [90]. Durch gemeinsames Nutzen von gleichen Mikrofonpositionen in verschiedenen Teilbändern kann diese Zahl auf 10 Mikrofone reduziert werden, ist damit aber immer noch zu groß für die meisten praktischen Anwendungen.

Aus diesen Gründen ist der Delay-and-Sum-Beamformer für die Unterdrückung breitbandiger kohärenter Störsignale ungeeignet. Die nachfolgend beschriebenen *Filter-and-Sum-Beamformer* stellen in dieser Hinsicht eine Verbesserung dar.

3.3.2 Filter-and-Sum-Beamformer

Filter-and-Sum-Beamformer bilden das Ausgangssignal durch Summation von gefilterten Sensorsignalen. Abbildung 3.7 zeigt die prinzipielle Struktur mit den Filtern $h_1(t) \dots h_M(t)$ bei einer Realisierung im Zeitbereich. Der Vorteil von Filter-and-Sum-Beamformern ist,

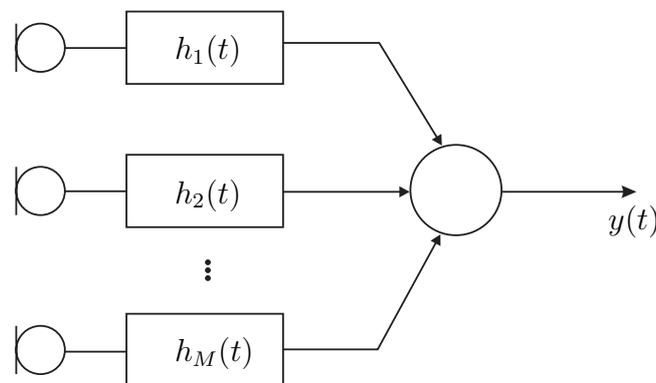


Abbildung 3.7: Allgemeine Struktur eines Filter and Sum Beamformers

daß in jedem Frequenzband ein separater Entwurf möglich wird, während bei Delay-and-Sum-Beamformern aufgrund von

$$\tau \circ \bullet e^{(-i\omega\tau)} \quad (3.41)$$

festen Beziehungen zwischen den Bändern bestehen. So läßt sich der im vorigen Kapitel angesprochene, auf Teilarrays basierende Entwurf, vorteilhaft umsetzen, da die Array-Gruppen nicht separat abgegriffen werden müssen, sondern mit den frequenzselektiven

Filtern realisiert werden können. Eine ausführliche Herleitung von frequenzinvarianten Beamformern dieser Art findet sich in [101] bzw. [102]. Allerdings gelten auch für diese Beamformer die Überlegungen aus dem vorherigen Abschnitt bezüglich der Mikrofonanzahl. So werden in [102] 25 Mikrofone für den Bereich zwischen 300 Hz und 3000 Hz eingesetzt, und in [88] werden insgesamt 16 Mikrofone verwendet. Für praktische Anwendungen ist die Anzahl an benötigten Mikrofonen damit zu hoch.

Die Verringerung der Anzahl der Mikrofone bei guter Richtwirkung ist nur durch Arrays mit superdirektiven Eigenschaften möglich. Die Theorie der Superdirektivität wurde bereits in den 50-er Jahren diskutiert (siehe z.B. [40]) und später erfolgreich auf Mikrofonarrays angewandt [108].

3.3.2.1 Superdirektiver Beamformer

Der Begriff *Superdirektivität* bezieht sich auf Beamformer, die trotz geringer Sensoranzahl eine gute Richtwirkung aufweisen. Voraussetzung dafür ist ein Sensorabstand von

$$d < \lambda/2. \quad (3.42)$$

Für die Herleitung des Superdirektiven Beamformers wird nachfolgend der Ansatz aus [14] in [17] benutzt. Er beruht auf der Minimierung der Ausgangssignalleistung des Beamformers im diffusen Rauschfeld bei gleichzeitiger Einhaltung einer Nebenbedingung. Diese ist gegeben als

$$\mathbf{w}^H(\omega) \mathbf{r}(\phi_{nutz}, \omega) = 1, \quad (3.43)$$

also als Forderung nach einer unverzerrten Übertragungsfunktion für Signale aus der Nutzsignalrichtung. Das Ausgangssignal wird in jedem Frequenzband aus der mit den Koeffizienten gewichteten Summe der Sensorsignale gebildet. Ohne die Abhängigkeit von ω weiter mitzuführen kann das Ausgangssignal geschrieben werden als

$$Y = \mathbf{w}^H \mathbf{X}. \quad (3.44)$$

Damit läßt sich die Ausgangssignalleistung in jedem Frequenzband wie folgt berechnen:

$$\Phi_{YY} = E[YY^*] \quad (3.45)$$

$$= \mathbf{w}^H E[\mathbf{X}\mathbf{X}^H] \mathbf{w} \quad (3.46)$$

$$= \mathbf{w}^H \Phi_{XX} \mathbf{w}, \quad (3.47)$$

wobei Φ_{XX} die Kreuzkorrelationsmatrix der Mikrofonssignale in jedem Frequenzband darstellt. Die Minimierung der Ausgangsleistung soll für Rauschsignale erfolgen, so daß sich mit Φ_{NN} als Korrelationsmatrix des Rauschens das Minimierungsproblem als

$$\min_{\mathbf{w}} \mathbf{w}^H \Phi_{NN} \mathbf{w} \quad (3.48)$$

formulieren läßt. Dieses komplexwertige Problem mit Nebenbedingung kann mit Hilfe der Lagrange-Multiplikator-Methode gelöst werden (siehe z.B. [93]) und spielt auch beim Entwurf adaptiver Beamformer eine zentrale Rolle. Für die Lösung der Minimierungsaufgabe wird eine Kostenfunktion mit eingearbeiteter Nebenbedingung aufgestellt

$$\mathcal{K}_{SDB} = \mathbf{w}^H \Phi_{NN} \mathbf{w} + \lambda (\mathbf{w}^H \mathbf{r} - 1). \quad (3.49)$$

Da die Nebenbedingung (3.43) reellwertig ist, kann auf die Verwendung komplexwertiger Lagrange-Multiplikatoren verzichtet werden [91]. Die Ableitung nach den konjugiert komplexen Koeffizienten [18] liefert

$$\frac{\partial \mathcal{K}_{SDB}}{\partial \mathbf{w}^H} = \Phi_{NN} \mathbf{w} + \lambda \mathbf{r}. \quad (3.50)$$

Durch Nullsetzen der Ableitung erhält man

$$\mathbf{w} = -\lambda \Phi_{NN}^{-1} \mathbf{r}. \quad (3.51)$$

Für die Bestimmung des Parameters λ muß diese Lösung in die Gleichung der Nebenbedingung (3.43) eingesetzt werden. Wegen $\Phi_{NN}^{-1} = \Phi_{NN}^{-H}$ gilt dann für die Nebenbedingung

$$1 = -\lambda \mathbf{r}^H \Phi_{NN}^{-1} \mathbf{r}. \quad (3.52)$$

Gleichung (3.52) kann nach λ umgestellt werden

$$\lambda = -\frac{1}{\mathbf{r}^H \Phi_{NN}^{-1} \mathbf{r}}, \quad (3.53)$$

womit sich schließlich der Koeffizientenvektor in jedem Frequenzband wie folgt berechnen läßt:

$$\mathbf{w}_{SDB} = \frac{\Phi_{NN}^{-1} \mathbf{r}}{\mathbf{r}^H \Phi_{NN}^{-1} \mathbf{r}}. \quad (3.54)$$

Nimmt man ein homogenes Rauschfeld an, bei dem die Korrelation unabhängig von der absoluten Sensorposition ist, läßt sich der optimale Koeffizientenvektor auch in Abhängigkeit

von der Kohärenzmatrix bestimmen. Die frequenzabhängige Kohärenzmatrix beinhaltet an der Stelle (i, j) die Kohärenzfunktionen zwischen den Sensoren i und j .

$$\mathbf{\Gamma}(\omega) = \begin{bmatrix} 1 & \gamma_{x_1x_2} & \gamma_{x_1x_3} & \cdots & \gamma_{x_1x_M} \\ \gamma_{x_2x_1} & 1 & \gamma_{x_2x_3} & \cdots & \gamma_{x_2x_M} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{x_Mx_1} & \gamma_{x_Mx_2} & \gamma_{x_Mx_3} & \cdots & 1 \end{bmatrix} \quad (3.55)$$

Da für verschiedene Rauschumgebungen die theoretischen Verläufe der jeweiligen Kohärenzfunktionen bekannt sind, lassen sich die Koeffizienten leicht an die akustische Situation anpassen. Bei Annahme eines diffusen Rauschfeldes mit si-förmiger Kohärenzfunktion ergeben sich die Koeffizienten des superdirektiven Beamformers als

$$\mathbf{w}_{SDB} = \frac{\mathbf{\Gamma}_{diffus}^{-1} \mathbf{r}}{\mathbf{r}^H \mathbf{\Gamma}_{diffus}^{-1} \mathbf{r}} \quad (3.56)$$

Als Beispiel für einen superdirektiven Entwurf dieser Art mit vier Mikrofonen im Abstand von 15 mm dient der Beamformer aus Abbildung 3.8. Allerdings läßt sich der Entwurf in dieser Form nicht in der Praxis einsetzen, da die Verstärkung unkorrelierten Rauschens bzw. die Norm der Koeffizienten bei der Herleitung nicht berücksichtigt wurden und beim gewählten Mikrofonabstand Werte über 150 dB annehmen.

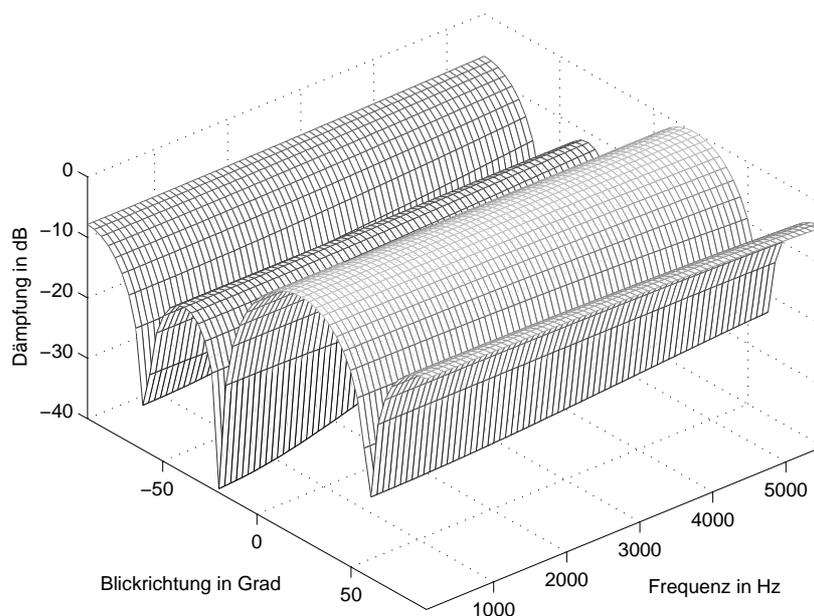


Abbildung 3.8: Superdirektiver Beamformer bei Nutzsignal aus 20° und vier Mikrofonen mit einem äquidistanten Abstand von $d = 15$ mm

Aus diesem Grund muß der Entwurf beschränkt und ein Kompromiß zwischen erreichbarer Unterdrückung korrelierter und Verstärkung unkorrelierter Signale gefunden werden. Dazu läßt sich die Kostenfunktion um einen Strafterm, der das Anwachsen des Betragsquadrates der Norm der Filterkoeffizienten auf einen festen Wert K_0 begrenzen soll, erweitern. Diese Nebenbedingung läßt sich als

$$\mathbf{w}^H \mathbf{w} = K_0 \quad (3.57)$$

formulieren, und sie kann mit Hilfe der Lagrange-Multiplikator-Methode in die Kostenfunktion integriert werden

$$\mathcal{K}_{SDB} = \mathbf{w}^H \Phi_{NN} \mathbf{w} + \lambda (\mathbf{w}^H \mathbf{r} - 1) + \mu (\mathbf{w}^H \mathbf{w} - K_0). \quad (3.58)$$

Es ist jedoch nicht möglich, eine geschlossene Lösung für μ anzugeben [40], [31]. Darum wird μ üblicherweise vorgegeben und die aus dieser Vorgabe resultierende Norm der Koeffizienten berechnet. Durch eine iterative Vorgehensweise läßt sich somit schließlich ein optimaler Wert für μ finden. Abbildung 3.9 zeigt das Beampattern eines auf diese Weise entworfenen superdirektiven Beamformers bei einer Beschränkung der Norm der Koeffizienten auf 10 dB.

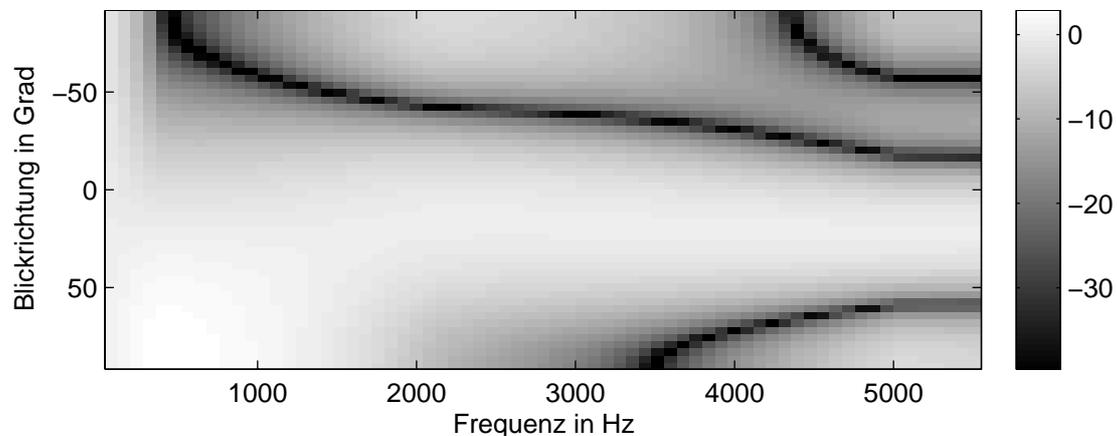


Abbildung 3.9: Superdirektiver Beamformer bei Nutzsignal aus 20° und vier Mikrofonen mit einem äquidistanten Abstand von $d = 15$ mm bei Entwurf mit begrenzter Koeffizientennorm von 10 dB

Es wird ersichtlich, daß die begrenzte Norm, deren Verlauf in Abbildung 3.10 dargestellt ist, zu einer signifikanten Verschlechterung der Richtwirkung im unteren Frequenzbereich führt. Eine ausführliche Analyse dieser Zusammenhänge und zu den Auswirkungen des Parameters μ ist in [31] zu finden.

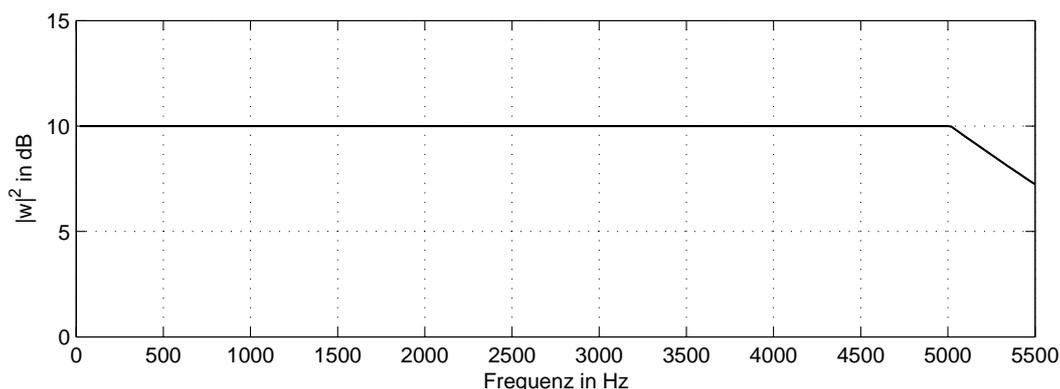


Abbildung 3.10: Norm der Koeffizienten des superdirektiven Beamformers bei Begrenzung auf 10 dB

3.3.2.2 Superdirektiver Entwurf für kohärente Störsignale

Da der im vorigen Abschnitt beschriebene superdirektive Beamformer für den Einsatz im diffusen Rauschfeld entworfen wurde, ist seine Richtwirkung für kohärente Störsignale nicht optimal. Stellt man sich z.B. eine akustische Situation vor, in der zwischen verschiedenen Sprechern unterschieden werden muß, wie es bei sprachgesteuerten Systemen im Kraftfahrzeug der Fall sein kann, muß das breitbandige, kohärente Störsprachsignal unterdrückt werden. Dabei sind die genauen Positionen von Fahrer und Beifahrer nur annähernd bekannt. Der Beamformer sollte also in der Lage sein, ganze Bereiche, aus denen möglicherweise das Störsignal kommen kann, auszublenden. Eine mögliche Bereichsunterscheidung zwischen Nutz- und Störsignal ist bei einem Mikrofonarray im Rückspiegel durch Annahme von $0^\circ \leq \phi_{\text{nutz}} \leq 90^\circ$ für die Fahrerposition und $-90^\circ \leq \phi_{\text{stör}} < 0^\circ$ für die Beifahrerposition gegeben.

Ein Design dieser Art entspricht dem klassischen Ansatz des nichtadaptiven Beamformer-Entwurfs für arbiträre Richtcharakteristika [97]. Dabei wird das gewünschte Richtverhalten als Zielfunktion vorgegeben und der bei der Approximation entstehende Fehler minimiert. Für die Formulierung dieses Problems als lineares Gleichungssystem muß der Vektor der Blickrichtungen

$$\boldsymbol{\phi} = [\phi_1, \phi_2, \dots, \phi_N]^T \quad (3.59)$$

und der Vektor der gewünschten Array-Antwort

$$\mathbf{b}(\boldsymbol{\phi}) = [\mathcal{V}(\phi_1), \mathcal{V}(\phi_2), \dots, \mathcal{V}(\phi_N)]^T \quad (3.60)$$

aufgestellt werden, wobei die einzelnen $\mathcal{V}(\phi_k)$ die gewünschte Verstärkung des Arrays bei einem Signal aus Richtung ϕ_k darstellen.

Die Arraygeometrie wird im Falle eines linearen, äquidistanten Arrays mit Sensorabstand d durch den Vektor der Mikrofonpositionen

$$\mathbf{d} = [0, d, 2d, \dots, (M-1)d]^T \quad (3.61)$$

beschrieben. Damit kann die aus den entsprechenden Phasenfaktoren bestehende Matrix der Einfallrichtungen, die von der Dimension $N \times M$ (Anzahl der Blickrichtungen \times Mikrofonanzahl) ist, wie folgt gebildet werden

$$\begin{aligned} \mathbf{A}(\omega) &= \exp\left(i\frac{\omega}{c} \cdot \boldsymbol{\phi} \mathbf{d}^T\right) \\ &= \begin{bmatrix} 1 & \exp(-i\omega\tau_2(\varphi_1)) & \exp(-i\omega\tau_3(\varphi_1)) & \dots & \exp(-i\omega\tau_M(\varphi_1)) \\ 1 & \exp(-i\omega\tau_2(\varphi_2)) & \exp(-i\omega\tau_3(\varphi_2)) & \dots & \exp(-i\omega\tau_M(\varphi_2)) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \exp(-i\omega\tau_2(\varphi_N)) & \exp(-i\omega\tau_3(\varphi_N)) & \dots & \exp(-i\omega\tau_M(\varphi_N)) \end{bmatrix}. \end{aligned} \quad (3.62)$$

Die Koeffizienten des Beamformers sind dann in jedem Frequenzband durch Lösung des Gleichungssystems

$$\mathbf{A}\mathbf{w} = \mathbf{b} \quad (3.64)$$

gegeben. Da die Anzahl der Blickrichtungen N meist größer als die Anzahl der Mikrofone M , die Matrix \mathbf{A} also nicht quadratisch ist, erhält man ein überbestimmtes Gleichungssystem ohne eindeutige Lösung. Das bedeutet, der Koeffizientenvektor muß im Sinne des kleinsten quadratischen Fehlers mit der Kostenfunktion

$$\mathcal{K} = |\mathbf{A}\mathbf{w} - \mathbf{b}|^2 \quad (3.65)$$

gefunden werden. Ein typischer Lösungsansatz ist, Gleichung (3.64) auf beiden Seiten mit \mathbf{A}^H zu multiplizieren, um die sogenannte Normalengleichung zu erhalten [67].

$$\mathbf{A}^H \mathbf{A} \mathbf{w} = \mathbf{A}^H \mathbf{b} \quad (3.66)$$

Das überbestimmte Gleichungssystem aus (3.64) ist damit auf ein quadratisches $M \times M$ System reduziert. Unter der Voraussetzung, daß $\mathbf{A}^H \mathbf{A}$ nicht singulär ist, berechnen sich die Koeffizienten zu

$$\mathbf{w}_{opt} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b}. \quad (3.67)$$

Dieser Ansatz entspricht einer Lösung mit Hilfe der Moore-Penrose-Pseudoinversen, die für \mathbf{A} mit vollem Rang wie folgt definiert ist:

$$\mathbf{A}^+ = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H. \quad (3.68)$$

Ist \mathbf{A} vom Rang $\varrho < M$, wie es bei kleinen Sensorabständen in den untereren Frequenzbändern vorkommen kann, muß \mathbf{A}^+ über eine Singulärwertzerlegung (siehe z.B. [83]) gefunden werden, da $(\mathbf{A}^H \mathbf{A})^{-1}$ nicht existiert. Aus $\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^H$ mit den orthogonalen Matrizen \mathbf{U} und \mathbf{V} sowie \mathbf{S} als Diagonalmatrix der Singulärwerte folgt:

$$\mathbf{A}^+ = \mathbf{V} (\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \mathbf{U}^H. \quad (3.69)$$

Für die Berechnung werden dabei nur die ersten ϱ , von Null verschiedenen Singulärwerte in \mathbf{S} benutzt, da andernfalls auch $(\mathbf{S}^H \mathbf{S})^{-1}$ nicht existiert. Damit kann die Lösung von Gleichung (3.65) schließlich wie folgt bestimmt werden:

$$\mathbf{w}_{opt} = \mathbf{A}^+ \mathbf{b}. \quad (3.70)$$

Ein Beispiel für einen auf diese Art und Weise berechneten Beamformer mit vier Mikrofonen und einem Mikrofonabstand von 1.5 cm zeigt Abbildung 3.11. Im Vergleich zum Delay-and-Sum-Beamformer ergibt sich auch für die unteren Frequenzbereiche eine sehr gute Störsignalunterdrückung von über 40 dB.

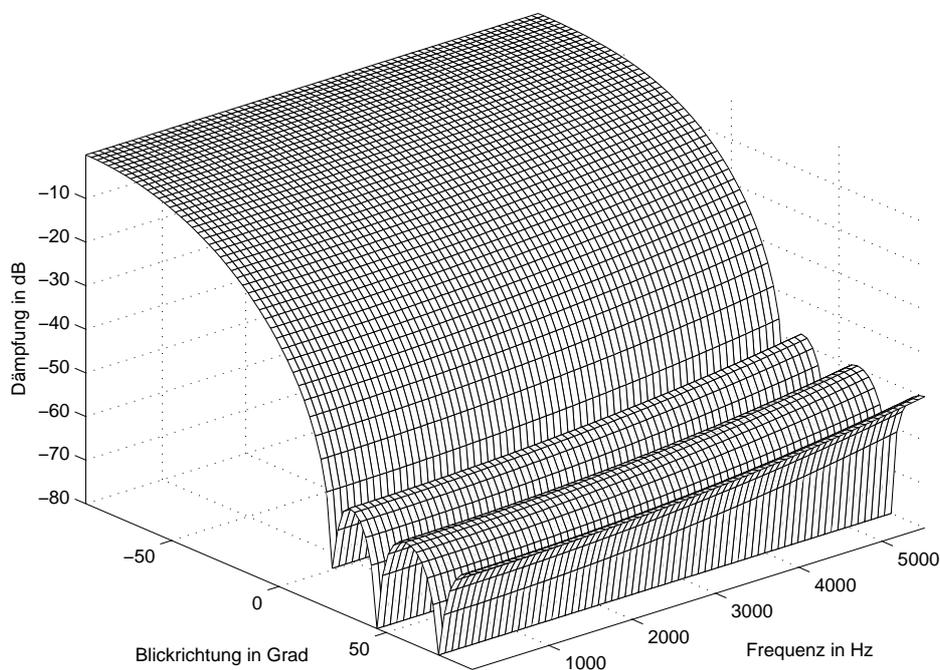


Abbildung 3.11: Beampattern eines superdirektiven Breitbandbeamformers im Frequenzbereich für vier Mikrofone mit einem äquidistanten Abstand von $d = 15$ mm

Obwohl die Lösung des Gleichungssystems (3.65) mit Hilfe der Pseudoinversen die Lösung mit dem kleinsten quadratischen Fehler und mit der kleinsten Norm der Koeffizienten darstellt, nimmt die Koeffizientennorm insbesondere in den unteren Frequenzbereichen noch sehr große Werte an. Wie sich in Abbildung 3.12 erkennen läßt, erreicht die Verstärkung unkorrelierten Rauschens Werte bis ca. 125 dB, wodurch der Beamformer in dieser Form unbrauchbar ist.

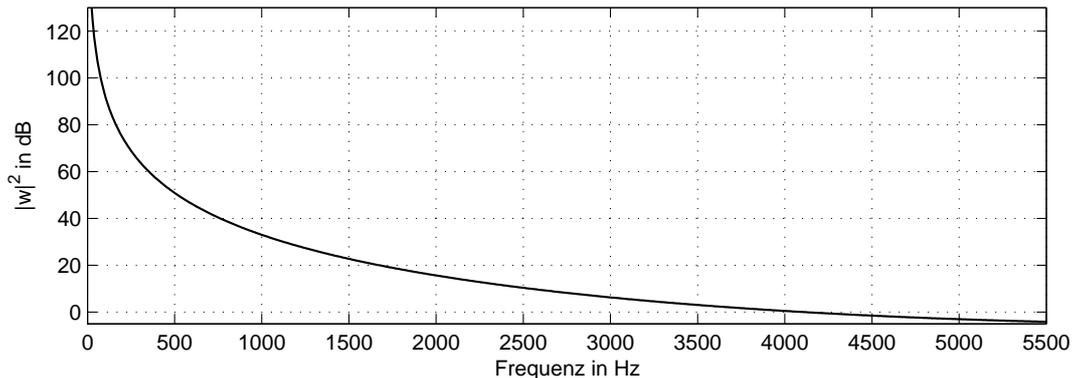


Abbildung 3.12: Norm der Koeffizienten des superdirektiven Beamformers aus Abbildung 3.11

Aus diesem Grund muß, vergleichbar mit der Vorgehensweise in Abschnitt 3.3.2.1, eine regularisierte Lösung benutzt werden. Die zwei verbreitetsten Regularisierungsverfahren sind die Tikhonov-Regularisierung und die Methode der begrenzten Singulärwertzerlegung (TSVD *Truncated Singular Value Decomposition*). Letztere führt in diesem Fall allerdings zu einer ungünstigeren Lösung, da im Beampattern Unstetigkeiten beim Wechsel der Anzahl der Singulärwerte auftreten.

Deshalb wird vorzugsweise die Tikhonov-Regularisierung benutzt. Die Kostenfunktion (3.65) erhält dabei einen zusätzlichen, gewichteten Strafterm, der das unkontrollierte Anwachsen der Norm der Koeffizienten verhindern soll

$$\mathcal{K}_{\text{reg}} = |\mathbf{A}\mathbf{w} - \mathbf{b}|^2 + \lambda^2 \mathbf{w}^H \mathbf{w} \quad (3.71)$$

$$= \mathbf{w}^H \mathbf{A}^H \mathbf{A} \mathbf{w} - \mathbf{w}^H \mathbf{A}^H \mathbf{b} - \mathbf{b}^H \mathbf{A} \mathbf{w} + \mathbf{b}^H \mathbf{b} + \lambda^2 \mathbf{w}^H \mathbf{w}. \quad (3.72)$$

Die Ableitung der regularisierten Kostenfunktion nach den konjugiert komplexen Koeffizienten [18] liefert

$$\frac{\partial \mathcal{K}_{\text{reg}}}{\partial \mathbf{w}^*} = \mathbf{A}^H \mathbf{A} \mathbf{w} - \mathbf{A}^H \mathbf{b} + \lambda^2 \mathbf{w}. \quad (3.73)$$

Durch Nullsetzen der Ableitung erhält man eine regularisierte Form der Normalgleichung

$$(\mathbf{A}^H \mathbf{A} + \lambda^2 \mathbf{I}) \mathbf{w}_{\text{reg}} = \mathbf{A}^H \mathbf{b}. \quad (3.74)$$

Um diese Gleichung zu lösen, muß der Parameter λ , der die Gewichtung des Strafterms darstellt, so gewählt werden, daß die gewünschte Ungleichungs-Nebenbedingung eingehalten wird. Diese ist als maximale Norm der Beamformerkoeffizienten vorgegeben, die, wie in Abschnitt 3.2.2.3 gezeigt, direkt mit der Verstärkung unkorrelierten Rauschens korrespondiert.

Für dieses Problem läßt sich wie bei dem Entwurfsverfahren im Abschnitt 3.3.2.1 keine geschlossene Lösung angeben, weshalb λ in jedem zu regularisierenden Frequenzband iterativ bestimmt werden muß. Durch die stufenweise Verkleinerung der Schrittweite kann λ hinreichend genau berechnet werden, so daß die Nebenbedingung bis auf eine vorzugebende Toleranz eingehalten wird.

Der in Abbildung 3.11 dargestellte Beamformer mit vier, im Abstand von 15 mm angeordneten Mikrofonen, läßt sich dann bei einer Vorgabe von $\mathbf{w}^H \mathbf{w} \leq 6$ dB auch in der Praxis realisieren. Abbildung 3.13 zeigt das zugehörige Beampattern mit der über einen großen Frequenzbereich guten Richtwirkung. Aufgrund des geringen Mikrofonabstandes ist jedoch auch bei diesem Entwurf der Gewinn für tiefe Frequenzen begrenzt, was sich in einer Dämpfung der Signale aus Nutzsignalrichtung äußert.

In Abbildung 3.14 ist der Verlauf der Norm der Koeffizienten dargestellt. Oberhalb von ca. 3000 Hz wird die Nebenbedingung automatisch von der nach Gleichung (3.70) gefundenen Lösung eingehalten, und die Norm der Koeffizienten nimmt kleinere Werte als 6 dB an.

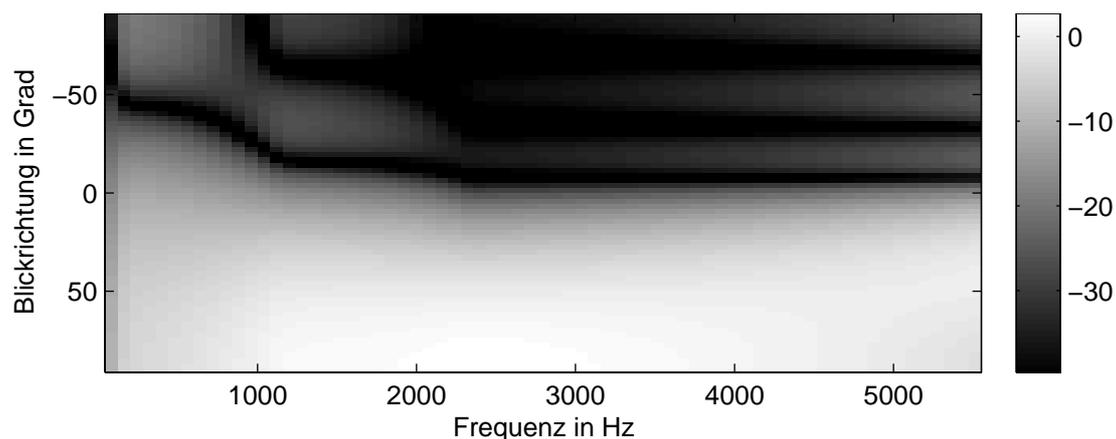


Abbildung 3.13: Beampattern des auf $\mathbf{w}^H \mathbf{w} \leq 6$ dB beschränkten superdirektiven Beamformers mit vier Mikrofonen im äquidistanten Abstand von $d = 15$ mm

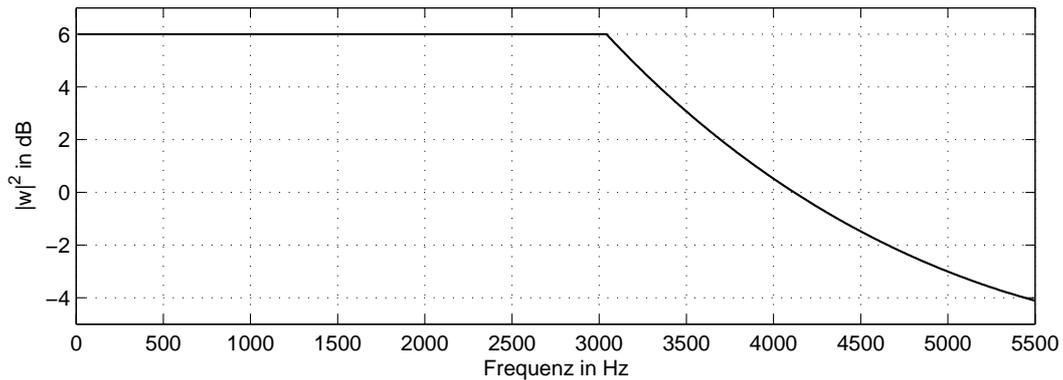


Abbildung 3.14: Norm der Koeffizienten des regularisierten superdirektiven Beamformers mit vier Mikrofonen im äquidistanten Abstand von $d = 15$ mm

3.4 Adaptive Beamformer

Adaptive Beamformer sind in der Lage, sich durch Nachregeln der Koeffizienten an die jeweilige akustische Situation anzupassen. Die Art und Weise des Nachregels entspricht dabei den aus der Theorie adaptiver Filter (siehe z.B. [44]) bekannten Verfahren wie LMS (*least mean square*) und RLS (*recursive least square*).

Die Unterscheidung der verschiedenen adaptiven Beamforming-Verfahren ist daher eher durch die jeweilige Struktur und durch das Optimierungskriterium gegeben. Nach [106] ist nach Vorarbeiten von Howell der adaptive *Multiple Sidelobe Canceller* von Applebaum [5] eine der ersten mehrkanaligen Strukturen. Bei diesem Verfahren wurden nutzsinalfreie Störreferenzen in unterschiedlicher Gewichtung mit einem Hauptkanal kombiniert. Die Adaption der Gewichte erfolgte nach dem Kriterium der minimalen Ausgangsleistung. Weitere Verfahren basierten nach [98], [64] z.B. auf der Verfügbarkeit einer Nutzsinalreferenz [105] oder auch auf der Maximierung des Signal-Rausch-Abstandes (SNR *signal to noise ratio*) [69]. Erst spätere Verfahren gingen zur Optimierung mit Nebenbedingungen über, die gegenwärtig den Stand der Technik darstellen.

3.4.1 Linear Constrained Minimum Variance Beamformer

Das erste Verfahren für adaptives Beamforming mit Nebenbedingungen stammt von Frost [38]. Die Struktur dieses Beamformers ist in Abbildung 3.15 auf der nächsten Seite dargestellt. Dabei handelt es sich um einen Beamformer für breitbandige Signale, der als Filter-and-Sum-Struktur mit variablen Koeffizienten ausgelegt ist.

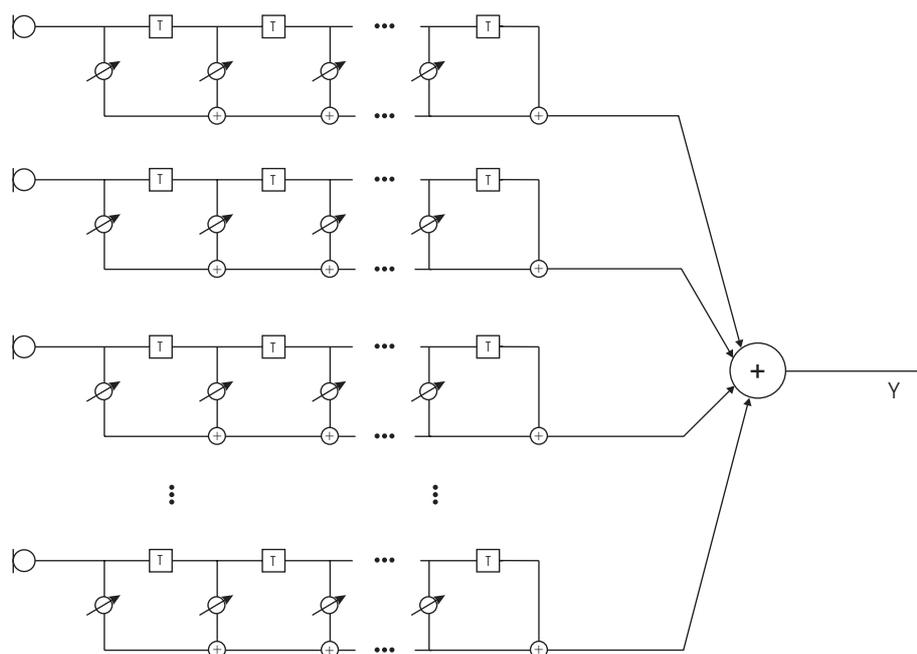


Abbildung 3.15: Struktur des Frost-Beamformers

Unter der Voraussetzung daß das Nutzsignal an allen Sensoren in gleicher Phasenlage zur Verfügung steht, kann die Filterung und Summation durch ein einzelnes äquivalentes FIR-Filter dargestellt werden. Die Koeffizienten dieses Filters bestehen dabei jeweils aus den aufsummierten Koeffizienten einer Spalte des Beamformers. Dieser Ansatz ermöglicht die konkrete Vorgabe einer FIR-Filterfunktion für das Nutzsignal, ohne die Anzahl der Freiheitsgrade übermäßig zu beschränken. Das zu optimierende Kriterium ist auch bei dieser Methode die Ausgangsleistung, und die Optimierung erfolgt mit einem LMS-Verfahren, wodurch die kumulative Ansammlung von Rundungsfehlern vermieden wird.

Eine im Vergleich zum Verfahren nach Frost andere Methode zur Einarbeitung einer Vorgabe für das Nutzsignal besteht in der Verwendung einer räumlichen Nebenbedingung, d.h. in der Vorgabe der Antwort des Beamformers auf ein Signal aus einer bestimmten Richtung. Mit Hilfe der Fernfeld-Parametrisierung kann die räumliche Nebenbedingung als

$$\mathbf{w}^H \mathbf{r} \stackrel{!}{=} g \quad (3.75)$$

formuliert werden, wobei g eine beliebige komplexe Konstante darstellt. Drückt man die Ausgangsleistung in Abhängigkeit von den Beamformerkoeffizienten aus, ergibt sich das

nachfolgende Minimierungsproblem

$$\min_{\mathbf{w}} \mathbf{w}^H \Phi_{\mathbf{X}\mathbf{X}} \mathbf{w}. \quad (3.76)$$

Gleichung (3.76) kann unter Einhaltung der Nebenbedingung (3.75) mit Hilfe der Lagrange-Multiplikator-Methode gelöst werden. Damit erhält man die Vorschrift für den LCMV-Beamformer (*Linearly Constrained Minimum Variance*)

$$\mathbf{w} = g^* \frac{\Phi_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{r}}{\mathbf{r}^H \Phi_{\mathbf{X}\mathbf{X}} \mathbf{r}}. \quad (3.77)$$

Wählt man $g = 1$, beschreibt Gleichung (3.77) einen MVDR-Beamformer (*Minimum Variance Distortionless Response*), bei dem das Signal aus Nutzsignalrichtung nicht gefiltert bzw. gestört wird (vgl. Abschnitt 3.3.2.1).

Allerdings ist die Bezeichnung von Gleichung (3.77) als LCMV-Beamformer in der Literatur nicht einheitlich. Während ein Großteil der Autoren ebenfalls diese Bezeichnung verwendet, findet man z.B. in [93] eine Formulierung des LCMV-Beamformers in Abhängigkeit von der Kovarianzmatrix des Rauschens. Die Form aus (3.77) wird dort als LCMP-Beamformer (*Linear Constrained Minimum Power*) bezeichnet. Meist wird jedoch zwischen diesen beiden Ansätzen gar nicht unterschieden, und es haben sich noch weitere Bezeichnungen, wie ML-Beamformer (*Maximum Likelihood*) [51] (üblicherweise bei Verwendung der Kovarianzmatrix des Rauschens) und Capon's Methode [77] etabliert.

Zusätzlich zu der einzelnen Nebenbedingung nach Gleichung (3.75) lassen sich weitere räumliche Vorgaben erzwingen. So ist es möglich, das Ausnullen einer bestimmten Richtung zu fordern. Eine Kombination der beiden beschriebenen Nebenbedingung läßt sich wie folgt formulieren [97]:

$$\begin{bmatrix} \mathbf{r}^H(\phi_1, \omega) \\ \mathbf{r}^H(\phi_2, \omega) \end{bmatrix} \mathbf{w} = \begin{bmatrix} g^* \\ 0 \end{bmatrix}. \quad (3.78)$$

Mehrere Nebenbedingungen werden üblicherweise in der Nebenbedingungsmatrix \mathbf{C} zusammengefaßt, so daß gilt

$$\mathbf{C}^H \mathbf{w} = \mathbf{f}, \quad (3.79)$$

wobei \mathbf{f} die Antwort des Beamformers bei Vorgabe der jeweiligen Nebenbedingung darstellt. Die Nebenbedingungen können dabei als Ableitungs-, Eigenvektor- oder punktweise Beschränkungen ausgeführt sein [97].

3.4.2 Generalized Sidelobe Canceller (GSC)

Eine alternative Struktur zur Realisierung eines LCMV-Beamformers, die inzwischen sehr weit verbreitet ist, wird in [42] beschrieben. Im Unterschied zur Struktur nach Frost ist der Beamformer in einen adaptiven und einen nicht adaptiven Signalpfad unterteilt.

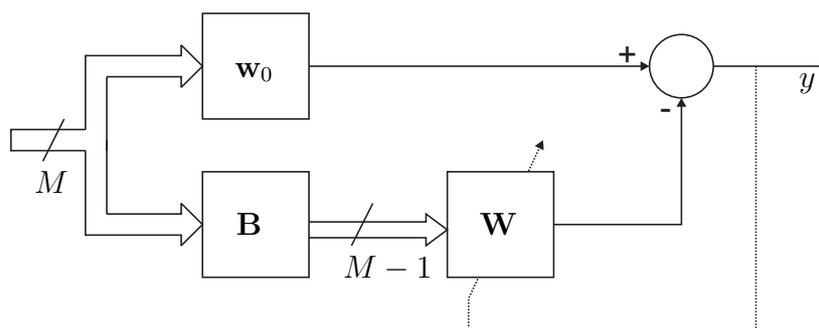


Abbildung 3.16: Struktur des Generalized Sidelobe Cancellers

Abbildung 3.16 zeigt ein Blockschaltbild dieser als *Generalized Sidelobe Canceller* (GSC) bezeichneten Struktur. Im oberen Pfad befindet sich der datenunabhängige M -kanalige Beamformer mit dem festen Koeffizientenvektor \mathbf{w}_0 . Der untere Pfad wird mit Hilfe der Blocking-Matrix \mathbf{B} von den Nutzsignalanteilen befreit, indem z.B. jeweils zwei benachbarte Kanäle voneinander subtrahiert werden [42]. Im Fall von vier Mikrofonen weist die Blocking-Matrix also die Form

$$\mathbf{B} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \quad (3.80)$$

auf. Dabei wird vorausgesetzt, daß die Nutzsignale an den Sensoren laufzeitkompensiert zur Verfügung stehen, das Array also auf die Nutzsignalquelle ausgerichtet ist. Erst dann ist theoretisch die vollständige Blockierung der Nutzsignalanteile möglich, da die Summe eines Zeilenvektors von \mathbf{B} immer gleich Null ist. Eine weitere Anforderung an die Blocking-Matrix ist, daß sie vollen Rang haben muß, die Zeilenvektoren also linear unabhängig voneinander sind.

Das nachfolgende mehrkanalige adaptive Filter mit dem Koeffizientensatz \mathbf{W} läßt sich aufgrund des ausgeblendeten Nutzsignals ohne Nebenbedingungen adaptieren, wodurch sich einfachere Adaptionsalgorithmen implementieren lassen. Zusätzlich kann der datenunabhängige Beamformerausgang z.B. für eine adaptive Nachfilterung genutzt werden [85].

Ihren Namen verdankt diese Struktur neben ihrer Ähnlichkeit zum *Multiple Sidelobe Canceller* der Tatsache, daß kohärente Störer aus Richtungen, bei denen der fest entworfene Beamformer Nebenmaxima aufweist, durch den adaptiven Teil stark reduziert werden. Es läßt sich weiterhin zeigen, daß bei Verwendung von konstanten Gewichten \mathbf{w}_0 entsprechend eines Delay-and-Sum-Beamformers der GSC identisch zum LCMV-Beamformer nach Frost ist [42].

4 Methoden der blinden Quellentrennung

Die mehrkanalige Signalerfassung ist in vielen verschiedenen Bereichen, wie z.B. der biomedizinischen Signalverarbeitung, der Sonar- und Radartechnik sowie auch in der Akustik wiederzufinden. Dabei steht man oft vor dem Problem, daß die mit Hilfe mehrerer Sensoren gemessenen Signale eigentlich Überlagerungen einzelner Quellsignale sind. In diesem Zusammenhang hat sich der Begriff *blinde Quellentrennung* bzw. BSS (*Blind Source Separation*) für Verfahren, die für die Rekonstruktion der Quellsignale nur sehr wenig Vorwissen über die Signale und deren Mischung nutzen, etabliert.

Eine große Klasse an Algorithmen dieser Art basiert auf der Annahme, daß die Quellsignale statistisch unabhängig voneinander sind. Durch die Bewertung der statistischen Unabhängigkeit, die *Independent Component Analysis* (ICA), läßt sich dann das Signalgemisch trennen bzw. einzelne oder mehrere Quellsignale extrahieren.

4.1 Signalmodelle

Zu den wenigen Annahmen, die im Vorfeld über eine Mischung zu treffen sind, zählt an erster Stelle die Frage nach der Art der Vermischung. Dabei wird prinzipiell zwischen instantaner, konvolutiver, linearer und nichtlinearer Mischung (bzw. Kombinationen davon) unterschieden. Die im Kontext dieser Arbeit wichtigsten Mischungsarten sind die instantane und die konvolute Mischung, die in den nachfolgenden Abschnitten erläutert werden.

4.1.1 Modell der linearen, instantanen Überlagerung

Das Modell der linearen, instantanen Überlagerung beschreibt einen Prozeß, bei dem die Quellsignale in Linearkombination von den Sensoren erfaßt werden. Die Mischung ist auf

die Augenblickswerte beschränkt und kann als Multiplikation mit skalaren Koeffizienten und anschließende Summation der Form

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + \dots + a_{1N}s_N(t) \quad (4.1)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + \dots + a_{2N}s_N(t) \quad (4.2)$$

⋮

$$x_M(t) = a_{M1}s_1(t) + a_{M2}s_2(t) + \dots + a_{MN}s_N(t) \quad (4.3)$$

dargestellt werden. Dabei stellen die $s_i(t)$ die Quellsignale, die a_{ij} die Mischungskoeffizienten und die x_i die Sensorsignale dar. Gleichung (4.1) bis (4.3) kann vorteilhaft in der Matrixschreibweise

$$\mathbf{x}(t) = \mathbf{A} \cdot \mathbf{s}(t) \quad (4.4)$$

dargestellt werden. Die sogenannte Mischungsmatrix \mathbf{A} ist allgemein von der Größe $M \times N$, mit M als Anzahl der Sensoren und N als Anzahl der Quellsignale. Sie enthält an der Stelle (i, j) den Gewichtungskoeffizienten a_{ij} , also den Faktor, mit dem Signal j an Sensor i vorliegt. Für die weiteren Ausführungen wird, sofern nicht anders bezeichnet, stets der quadratische 2×2 Fall nach Abbildung 4.1 betrachtet, bei dem die Anzahl an Quellen gleich der Anzahl an Sensoren gleich zwei ist.

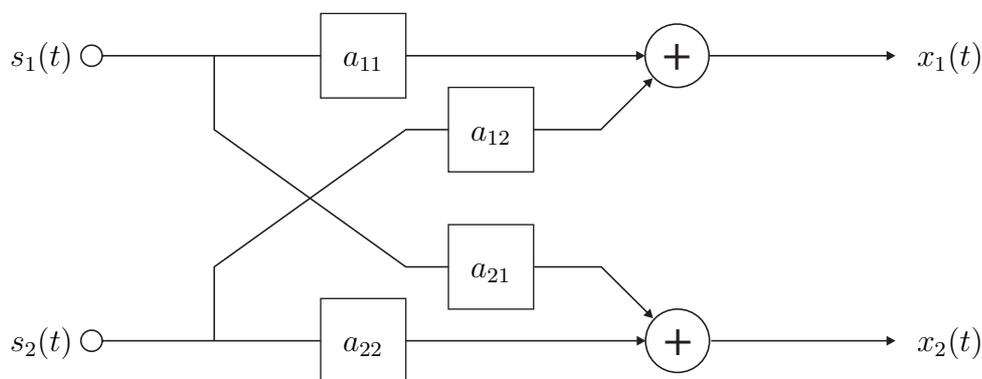


Abbildung 4.1: Modell der instantanen Mischung zweier Signale

Vorausgesetzt daß \mathbf{A} vollen Rang hat, ist eine Lösung des Separationsproblems durch die Anwendung der inversen Matrix $\mathbf{W} = \mathbf{A}^{-1}$ gegeben. Die entmischten Ausgangssignale werden durch Multiplikation der Inversen mit den Sensorsignalen gebildet

$$\mathbf{y}(t) = \mathbf{W} \cdot \mathbf{x}(t). \quad (4.5)$$

Bei exakt bestimmter, invertierbarer Mischungsmatrix gilt

$$\mathbf{y}(t) = \mathbf{W} \cdot \mathbf{A}\mathbf{s}(t) = \mathbf{I} \cdot \mathbf{s}(t) = \mathbf{s}(t), \quad (4.6)$$

wobei \mathbf{I} die Einheitsmatrix darstellt. Bei einem überbestimmten Gleichungssystem mit mehr Sensoren als Quellen ($M > N$) läßt sich die Moore-Penrose-Pseudoinverse von \mathbf{A} zur Entmischung benutzen [1].

Für alle ICA-basierten Verfahren gelten die Einschränkungen, daß die Quellsignale nur bis auf eine mögliche Skalierung und Vertauschung (Permutation) bestimmt werden können, da skalierte und permutierte Quellsignale ebenfalls statistisch unabhängig voneinander sind. Mathematisch werden Skalierung als Diagonalmatrix \mathbf{D} und Vertauschung als Permutationsmatrix \mathbf{P} berücksichtigt, so daß sich die entmischten Ausgangssignale zu

$$\mathbf{y}(t) = \mathbf{PD} \cdot \mathbf{s}(t) \quad (4.7)$$

ergeben. Ein weit verbreiteter Ansatz zur Bestimmung von \mathbf{D} ist, die Skalierung durch Normierung der Eingangssignale auf Einheitsvarianz oder auch durch die Annahme eines normierten Mischungsmodells mit Einsen auf den Diagonalpositionen der Mischungsmatrix festzulegen. Bei Annahme des normierten Mischungsmodells ergibt sich gleichzeitig eine feste Zuordnung der Quellsignale zu den Sensoren

4.1.2 Modell der linearen, konvolutiven Überlagerung

Neben der instantanen Mischung spielt das konvolute Mischungsmodell nach Abbildung 4.2 auf der nächsten Seite bei vielen realen Anwendungen eine wichtige Rolle. Hier werden die Sensorsignale aus der Überlagerung gefilterter Quellsignale gebildet, wie es z.B. bei Mikrofonaufnahmen der Fall ist. Mit $h_{ij}(t)$ als Impulsantworten der Mischungsfilter ergeben sich die Sensorsignale im 2×2 Fall zu

$$x_1(t) = h_{11}(t) * s_1(t) + h_{12}(t) * s_2(t) \quad (4.8)$$

$$x_2(t) = h_{21}(t) * s_1(t) + h_{22}(t) * s_2(t). \quad (4.9)$$

Auch dieses Gleichungssystem kann als

$$\mathbf{x}(t) = \mathbf{A}(t) * \mathbf{s}(t) \quad (4.10)$$

in Matrixschreibweise dargestellt werden (siehe z.B. [58]), was insbesondere dann günstig ist, wenn im Zeitbereich arbeitende ICA-Verfahren zum Einsatz kommen. Die Mischungsmatrix $\mathbf{A}(t)$ enthält an der Stelle (i, j) keinen skalaren Wert sondern die Impulsantwort des Filters $h_{ij}(t)$ von Quelle j zum Sensor i .

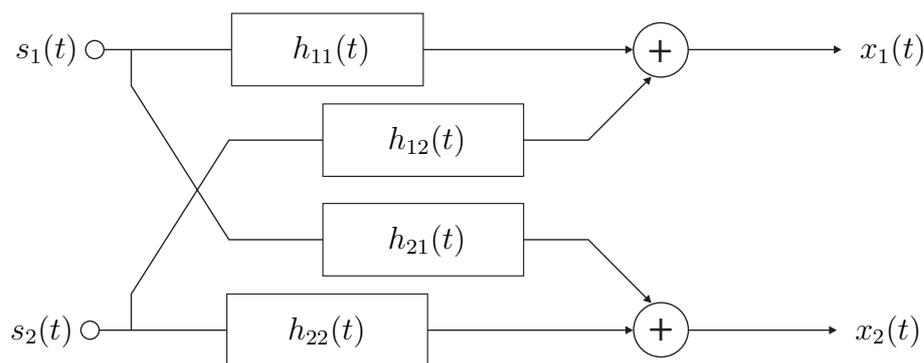


Abbildung 4.2: Modell der konvolutiven Mischung zweier Signale

Eine andere weit verbreitete Methode der Darstellung und Lösung des konvolutiven Mischungsproblems ist durch die Formulierung von Gleichung (4.8) und (4.9) im Frequenzbereich gegeben. Nach Anwendung der Kurzzeit-Fouriertransformation entsteht in jedem Frequenzband ein komplexwertiges, instantanes Mischungsproblem

$$X_1(\omega, \tilde{t}) = H_{11}(\omega) \cdot S_1(\omega, \tilde{t}) + H_{12}(\omega) \cdot S_2(\omega, \tilde{t}) \quad (4.11)$$

$$X_2(\omega, \tilde{t}) = H_{21}(\omega) \cdot S_1(\omega, \tilde{t}) + H_{22}(\omega) \cdot S_2(\omega, \tilde{t}). \quad (4.12)$$

Die Schreibweise in Matrixform

$$\mathbf{X}(\omega, \tilde{t}) = \mathbf{A}(\omega) \cdot \mathbf{S}(\omega, \tilde{t}) \quad (4.13)$$

ist äquivalent zu Gleichung (4.4). Allerdings enthält die komplexe Mischungsmatrix in jedem Frequenzband ω an der Stelle (i, j) die Werte der jeweiligen Übertragungsfunktion $H_{ij}(\omega)$. Es müssen dann für die Lösung des konvolutiven Mischungsproblems im Frequenzbereich insgesamt $\frac{L}{2}$ Matrizen $\mathbf{A}(\omega)$ bestimmt werden, wenn die Blocklänge L gleich der Anzahl der Frequenzbänder ist und der konjugiert komplexe Teil des Spektrums bei der Berechnung nicht berücksichtigt wird.

Die frequenzbandweise Bestimmung einzelner Mischungsmatrizen ermöglicht die Verwendung beliebiger instantaner, komplexwertiger ICA-Verfahren und liefert deshalb bei geringerem Rechenaufwand sehr gute Ergebnisse. Nachteilig an dieser Vorgehensweise ist, daß aufgrund der möglichen Permutationen keine eindeutige Zuordnung der Frequenzbänder gegeben ist. Ohne entsprechende Permutationskorrektur (siehe Abschnitt 4.3) ist deshalb keine oder nur eine sehr schlechte Signaltrennung möglich.

Ebenso wie das Permutationsproblem, müssen auch die möglichen Skalierungen der einzelnen Frequenzbänder korrigiert werden, die sich andernfalls als willkürliche Filterung

auswirken. Aus diesem Grund geht man häufig von einem Modell mit normierten Mischungsmatrizen aus. Gleichung (4.11) und (4.12) lassen sich dann wie folgt umformulieren

$$X_1(\omega, \tilde{t}) = S_1(\omega, \tilde{t}) + \tilde{H}_{12}(\omega) \cdot S_2(\omega, \tilde{t}) \quad (4.14)$$

$$X_2(\omega, \tilde{t}) = \tilde{H}_{21}(\omega) \cdot S_1(\omega, \tilde{t}) + S_2(\omega, \tilde{t}). \quad (4.15)$$

Die Mischungsmatrix hat also in jedem Frequenzband die Form

$$\mathbf{A}(\omega) = \begin{bmatrix} 1 & \tilde{H}_{12}(\omega) \\ \tilde{H}_{21}(\omega) & 1 \end{bmatrix}, \quad (4.16)$$

mit $\tilde{H}_{12}(\omega) = \frac{H_{12}(\omega)}{H_{22}(\omega)}$ und $\tilde{H}_{21}(\omega) = \frac{H_{21}(\omega)}{H_{11}(\omega)}$. Die Normierung ist, physikalisch gesehen, in jeder Zeile der Mischungsmatrix äquivalent zur Annahme eines Referenzsensors. Bei der zweikanaligen Schallerfassung mit Mikrofonen beschreibt das Filter $\tilde{H}_{12}(\omega)$, welcher Veränderung Signal zwei bei der Messung an Mikrofon eins im Vergleich zur Messung an Mikrofon zwei unterliegt. Analog dazu beschreibt Filter $\tilde{H}_{21}(\omega)$ die Veränderungen von Quellsignal eins an Mikrofon zwei im Vergleich zu Mikrofon eins. Allerdings können die beiden Filter nicht mit der tatsächlichen Übertragungsfunktion zwischen den Mikrofonen gleichgesetzt werden.

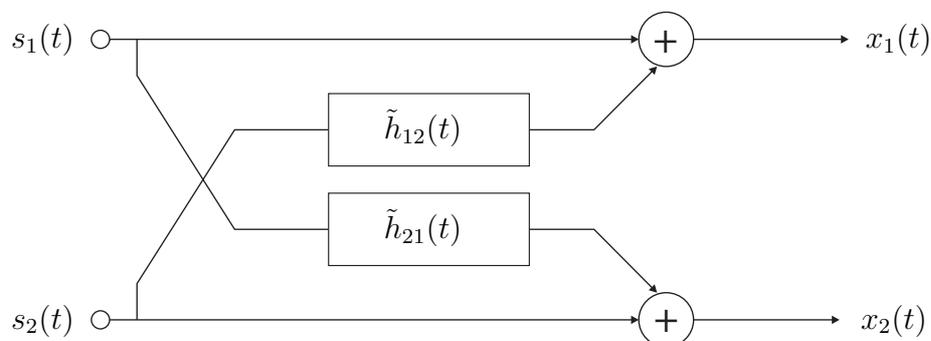


Abbildung 4.3: Modell der normierten, konvolutiven Mischung zweier Signale

Weiterhin ist offensichtlich, daß durch die Annahme eines normierten Mischungsmodells mit Referenzsensoren keine Raumübertragungsfunktionen kompensiert werden können. Das heißt, das Quellsignal $s_i(t)$ kann bestenfalls so rekonstruiert werden, wie es an Mikrofon i meßbar ist. Deshalb ist mit diesem Modell zwar die Trennung konvolutiver Mischungen, aber keine Entfaltung der Meßsignale möglich. Dies stellt jedoch keine Einschränkung dar, weil das Kriterium der statistischen Unabhängigkeit ohnehin keine Informationen über eine eventuelle Filterung der Signale liefert.

Für die Entmischung muß in jedem Frequenzband die Inverse der Mischungsmatrix bestimmt werden. Im 2×2 Fall ist sie gegeben durch

$$\mathbf{W}(\omega) = \mathbf{A}^{-1}(\omega) = \frac{1}{\tilde{H}_{12}(\omega)\tilde{H}_{21}(\omega) - 1} \cdot \begin{bmatrix} -1 & \tilde{H}_{12}(\omega) \\ \tilde{H}_{21}(\omega) & -1 \end{bmatrix}. \quad (4.17)$$

Die Entmischung kann nach erfolgreicher Permutationskorrektur im Frequenzbereich oder durch Rücktransformation und Filterung im Zeitbereich erfolgen.

4.1.2.1 Entmischungsstrukturen

Oft wird in der Literatur zwischen *Feed-Forward*- und *Feed-Back*-Strukturen zur Trennung gefalteter Mischungen unterschieden. Im 2×2 Fall besteht die *Feed-Forward*-Struktur aus vier nicht rückgekoppelten Filtern, ist also identisch zum Modell der konvolutiven Mischung nach Abbildung 4.2. Mit dieser Struktur ließe sich zusätzlich zur Signaltrennung auch eine Entfaltung realisieren, sofern eine geeignete Kostenfunktion zur Verfügung stünde. Die Filter können dabei direkt aus der Entmischungsmatrix abgelesen werden. Durch zusätzliche Allpaß-Filterung kann der akausale Anteil von nicht minimalphasigen Mischungssystemen kompensiert werden [61].

Bei Annahme einer normierten Mischungsmatrix, die die Entfaltung von vornherein ausschließt, kann die Entmischung aufwandsgünstiger in der rückgekoppelten *Feed-Back*-Struktur nach Abbildung 4.4 realisiert werden. Vorteilhaft ist dabei, daß nur zwei Filter bestimmt und implementiert werden müssen.

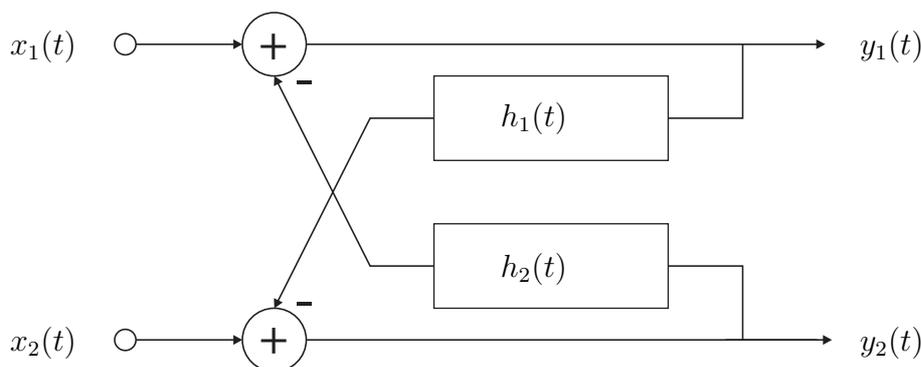


Abbildung 4.4: Feed-Back-Struktur zur Entmischung

Die beiden Ausgangssignale sind im z -Bereich gegeben durch

$$Y_1(z) = X_1(z) - H_2(z)Y_2(z) \quad (4.18)$$

$$Y_2(z) = X_2(z) - H_1(z)Y_1(z). \quad (4.19)$$

Durch Einsetzen von (4.19) in (4.18) kann die Abhängigkeit von einem Ausgangssignal eliminiert werden, so daß sich für $Y_1(z)$ ergibt

$$\begin{aligned} Y_1(z) &= X_1(z) - H_2(z)[X_2(z) - H_1(z)Y_1(z)] \\ &= X_1(z) - H_2(z)X_2(z) + H_2(z)H_1(z)Y_1(z) \\ &= \frac{X_1(z) - H_2(z)X_2(z)}{1 - H_1(z)H_2(z)}. \end{aligned} \quad (4.20)$$

Analog läßt sich für $Y_2(z)$ herleiten

$$Y_2(z) = \frac{X_2(z) - H_1(z)X_1(z)}{1 - H_1(z)H_2(z)}. \quad (4.21)$$

Aufgrund des normierten Mischungsmodells sind die Sensorsignale im z -Bereich, gemäß Gleichung (4.14) und (4.15) auf Seite 47, gegeben als

$$X_1(z) = S_1(z) + \tilde{H}_{12}(z)S_2(z) \quad (4.22)$$

$$X_2(z) = \tilde{H}_{21}(z)S_1(z) + S_2(z). \quad (4.23)$$

Eingesetzt in 4.20 kann $Y_1(z)$ in Abhängigkeit von den Quellsignalen formuliert werden

$$\begin{aligned} Y_1(z) &= \frac{S_1(z) + \tilde{H}_{12}(z)S_2(z) - H_2(z)[\tilde{H}_{21}(z)S_1(z) + S_2(z)]}{1 - H_1(z)H_2(z)} \\ &= S_1(z) \left[\frac{1 - H_2(z)\tilde{H}_{21}(z)}{1 - H_1(z)H_2(z)} \right] + S_2(z) \left[\frac{\tilde{H}_{12}(z) - H_2(z)}{1 - H_1(z)H_2(z)} \right]. \end{aligned} \quad (4.24)$$

Für die Entmischung soll $Y_1(z) = S_1(z)$ gelten. Daraus können die Bedingungen

$$H_1(z) = \tilde{H}_{21}(z) \quad (4.25)$$

$$H_2(z) = \tilde{H}_{12}(z) \quad (4.26)$$

abgeleitet werden. Die gleiche Vorschrift ergibt sich bei Betrachtung von $Y_2(z)$. Folglich lassen sich die rekursiven Filter direkt aus der normierten Mischungsmatrix ablesen, sofern diese korrekt bestimmt wurde. Die Invertierung dieser Matrix ist damit nicht mehr notwendig, und die damit verbundenen Stabilitätsprobleme entfallen.

Lediglich eventuell auftretende Nullstellen in den Übertragungsfunktionen $H_{11}(\omega)$ und $H_{22}(\omega)$ könnten aufgrund der Division durch Null in den normierten Filtern zu Problemen führen. Diese Situation stellt jedoch einen in der Praxis der Schallausbreitung quasi nicht vorkommenden Sonderfall dar und kann deswegen vernachlässigt werden.

Weiterhin läßt sich feststellen, daß die Stabilität des rekursiven Systems nicht, wie bei inversen Filtern, von der Minimalphasigkeit der Mischungsmatrix abhängt, die für reale Raumimpulsantworten meist nicht gegeben ist [61]. Sie wird statt dessen durch die Determinante der Mischungsmatrix bestimmt. Solange die Mischungsmatrix vollen Rang hat, ist das rekursive System genauso stabil wie die Feed-Forward-Struktur.

4.2 Blinde Quellentrennung instantaner Mischungen

Die Analyse der statistischen Unabhängigkeit (ICA *Independent Component Analysis*) ist das Fundament vieler Methoden der blinden Quellentrennung. Das Konzept statistisch unabhängiger Quellsignale kann bis in die 50-er Jahre des 20. Jahrhunderts zurückverfolgt werden [1]. Allerdings ist das breite Interesse an dieser Thematik erst in den späten 80-er Jahren durch die kontinuierlich wachsende Anzahl an Anwendungen entstanden. Bis zum heutigen Tage haben sich eine Vielzahl von Verfahren etabliert, die auf unterschiedlichen Kriterien zur Bewertung der statistischen Unabhängigkeit beruhen.

4.2.1 Definition der statistischen Unabhängigkeit

Als statistisch unabhängig werden zwei Zufallsvariablen genau dann bezeichnet, wenn ihre Verbundverteilungsdichte gleich dem Produkt der Randdichteverteilungen ist [47]. Folglich muß dann gelten

$$p_{\mathbf{x}}(\mathbf{x}) = p_{\mathbf{x}}(x_1, x_2) = p_{x_1}(x_1) \cdot p_{x_2}(x_2). \quad (4.27)$$

Ein Maß für die Gleichheit zweier Wahrscheinlichkeitsdichten ist die Kullback-Leibler-Divergenz, die als

$$D(p_y||p_x) = \int p_y(y) \log \frac{p_y(y)}{p_x(x)} dy \quad (4.28)$$

definiert ist [27]. Sie ist ein asymmetrisches Maß für den Abstand zwischen den Wahrscheinlichkeitsdichteverteilungen und ist invariant bezüglich Vertauschung und Skalierung [43]. Es läßt sich zeigen, daß die Ungleichung

$$D(p_y||p_x) \geq 0 \quad (4.29)$$

gelten muß, mit Gleichheit für $p_y(y) = p_x(x)$. Die Kullback-Leibler-Divergenz kann deshalb zur Bewertung der statistischen Unabhängigkeit genutzt werden. Angewendet auf (4.27) ergibt sich

$$D(p_{\mathbf{x}}(\mathbf{x})||p_{x_1}(x_1)p_{x_2}(x_2)) = \int p_{\mathbf{x}}(\mathbf{x}) \log \frac{p_{\mathbf{x}}(\mathbf{x})}{p_{x_1}(x_1)p_{x_2}(x_2)} d\mathbf{x}. \quad (4.30)$$

Für $D(p_{\mathbf{x}}(\mathbf{x})||p_{x_1}(x_1)p_{x_2}(x_2)) = 0$ sind die beiden Zufallsvariablen x_1 und x_2 statistisch unabhängig voneinander.

Die vollständige Auswertung der Verteilungsdichte nach Gleichung (4.30) ist für reale Anwendungen jedoch nicht praktikabel, da sich Verteilungsdichten nur mit großem Aufwand schätzen lassen. Aus diesem Grund versucht man, die Zufallsvariablen bzw. deren

statistische Unabhängigkeit anhand ausgewählter Parameter zu beschreiben. Klassische ICA-Verfahren stützen sich dabei vor allem auf statistische Kenngrößen höherer Ordnung, da die Unkorreliertheit der Variablen nicht ausreichend ist.

4.2.1.1 Dekorrelation als Vorverarbeitung

Eine bei ICA Algorithmen oft verwendete Methode der Vorverarbeitung ist das sogenannte *Sphering* bzw. *Prewhitening*; die Dekorrelation und Varianz-Normierung der Sensorsignale. Diese Methode ist sehr eng mit der Hauptkomponentenanalyse, auch als PCA (*Principal Component Analysis*) bezeichnet, verwandt. Ziel der PCA ist es, durch eine geeignete Transformation eine neue Darstellung der Signale in einem orthogonalen Koordinatensystem zu finden. Das heißt, die Kovarianzmatrix der dekorrelierten Daten ist diagonal.

Für die Herleitung der Hauptkomponentenanalyse (siehe z.B. [53]) wird der Signalvektor als gewichtete Überlagerung der orthogonalen Basisvektoren dargestellt. Bei reduzierter Anzahl an Basisvektoren entsteht ein Approximationsfehler, der unter Einhaltung der Orthogonalitätsbedingung minimiert wird. Mit Hilfe der Lagrange-Multiplikator-Methode ergibt sich die Lösung als Eigenwertproblem der Kovarianzmatrix. Die Signalapproximation erfolgt durch die zu den größten Eigenwerten zugehörigen Eigenvektoren. Damit ist eine Abschätzung der Anzahl der in den Sensordaten enthaltenen Signale [96], eine Dimensionsreduktion sowie die Schätzung der Rauschleistung möglich [91].

Analog dazu kann das Sphering als PCA der Sensordaten (ohne Dimensionsreduktion) mit anschließender Normierung der Signalvarianzen auf Eins betrachtet werden. Die Eigenvektorzerlegung der Kovarianzmatrix \mathbf{C} in

$$\mathbf{C} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T, \quad (4.31)$$

mit \mathbf{V} als orthogonaler Matrix der Eigenvektoren und $\mathbf{\Lambda}$ als Diagonalmatrix der zugehörigen Eigenwerte, führt zur Berechnung der Quadratwurzel der Kovarianzmatrix als

$$\mathbf{C}^{\frac{1}{2}} = \mathbf{V}\mathbf{\Lambda}^{\frac{1}{2}}. \quad (4.32)$$

Die Spheringmatrix \mathbf{M} ist als Inverse der Matrixwurzel gegeben mit

$$\mathbf{M} = \mathbf{C}^{-\frac{1}{2}}. \quad (4.33)$$

Durch Multiplikation mit den Sensordaten \mathbf{x} entstehen die transformierten Daten

$$\mathbf{x}_s = \mathbf{M}\mathbf{x} = \mathbf{C}^{-\frac{1}{2}}\mathbf{x}, \quad (4.34)$$

deren Kovarianzmatrix

$$\mathbf{C}_{x_s} = E[\mathbf{x}\mathbf{x}^T] = \mathbf{C}^{-\frac{1}{2}} E[\mathbf{x}\mathbf{x}^T] \mathbf{C}^{-\frac{1}{2}T} = \mathbf{C}^{-\frac{1}{2}} \mathbf{C} \mathbf{C}^{-\frac{1}{2}T} = \mathbf{I} \quad (4.35)$$

eine Einheitsmatrix ist [53].

Ausgehend vom Mischungsmodell nach Gleichung (4.4) auf Seite 44 lassen sich die transformierten Signale \mathbf{x}_s aus Gleichung (4.34) darstellen als

$$\mathbf{x}_s = \mathbf{M}\mathbf{x} = \mathbf{M}\mathbf{A}\mathbf{s} = \mathbf{O}\mathbf{s}, \quad \text{mit } \mathbf{O} = \mathbf{M}\mathbf{A}. \quad (4.36)$$

Durch das Prewhitening ist sichergestellt, daß gilt

$$\mathbf{C}_{x_s} = E[\mathbf{x}_s\mathbf{x}_s^T] = \mathbf{O} E[\mathbf{s}\mathbf{s}^T] \mathbf{O}^T = \mathbf{I}. \quad (4.37)$$

Aufgrund der per Definition angenommenen statistischen Unabhängigkeit der Quellsignale sowie aufgrund der Normierung der Varianz ist auch die Kovarianzmatrix der Quellsignale eine Einheitsmatrix

$$E[\mathbf{s}\mathbf{s}^T] = \mathbf{I}, \quad (4.38)$$

so daß

$$\mathbf{O}\mathbf{O}^T = \mathbf{I} \quad (4.39)$$

gelten muß [53]. Folglich ist die Mischungs- bzw. Entmischungsmatrix der transformierten Daten orthogonal. Dadurch vereinfacht sich die Suche nach der Entmischungsmatrix erheblich, so daß viele ICA Verfahren auf dieser Form der Vorverarbeitung mit anschließender Suche nach einer orthogonalen Entmischungsmatrix basieren.

4.2.2 Prinzipien der Informationstheorie

Die historische Entwicklung der ICA-Verfahren ist eng mit Ansätzen aus dem Bereich der Neuronalen Netze, wie der Redundanzreduktion [6] oder der Maximierung der übertragenen Information [62] verbunden [53]. Auf dieser Grundlage entstanden eine Vielzahl von informationstheoretisch motivierten Verfahren, deren Konzepte nachfolgend kurz vorgestellt werden sollen.

Eine wichtige Kenngröße der Informationstheorie ist die Shannonsche Entropie $H(x)$, die ein Maß für den Informationsgehalt einer Datenquelle darstellt [50]

$$H(x) = - \sum_i P(x_i) \log P(x_i). \quad (4.40)$$

Für wertkontinuierliche Quellen muß mit der differentiellen Entropie gearbeitet werden, da die Entropie beim Grenzübergang $\Delta x \rightarrow 0$ divergiert. Sie ist definiert als

$$H(x) = - \int_{-\infty}^{\infty} p_x(x) \log(p_x(x)) dx. \quad (4.41)$$

Für Zufallsvariablen mit begrenzter Amplitude weisen gleichverteilte Variablen die größte Entropie auf. Hingegen gilt für Zufallsvariablen mit fester Varianz, daß die Entropie durch die Gaußverteilung maximiert wird. Als Maß für die Abweichung einer Zufallsvariable mit gegebener Varianz von der Gaußverteilung läßt sich die nachfolgend für multivariate Zufallsvariablen definierte Negentropie

$$\mathcal{N}(\mathbf{x}) = H_{\text{Gauß}} - H(\mathbf{x}) \quad (4.42)$$

verstehen. Die Negentropie nimmt keine negativen Werte an und hat weiterhin die Eigenschaft, sowohl skalierungsinvariant als auch invariant gegenüber orthogonalen Transformationen zu sein [26].

Die in Abschnitt 4.2.1 eingeführte Definition der statistischen Unabhängigkeit als Kullback-Leibler-Divergenz zwischen dem Produkt der Randdichteverteilungen und der Verbundverteilung ist äquivalent zur gemeinsamen Information (*Mutual Information*)

$$I(x_1, \dots, x_N) = D(p_{\mathbf{x}} || \prod p_{x_n}), \quad (4.43)$$

deren Reduktion folglich zur statistischen Unabhängigkeit führt [26, 2]. Aus der Definition der Kullback-Leibler-Divergenz nach (4.28) folgt:

$$I(x_1, \dots, x_N) = \int p_{\mathbf{x}}(\mathbf{x}) \log \frac{p_{\mathbf{x}}(\mathbf{x})}{\prod p_{x_n}(x_n)} d\mathbf{x} \quad (4.44)$$

$$= \int p_{\mathbf{x}}(\mathbf{x}) \log p_{\mathbf{x}} - \int p_{\mathbf{x}}(\mathbf{x}) \log \prod p_{x_n}(x_n) \quad (4.45)$$

$$= \sum_{n=1}^N H(x_n) - H(\mathbf{x}). \quad (4.46)$$

Die gemeinsame Information einer multivariaten Zufallsvariable läßt sich also als Summe der Einzelentropien abzüglich der gesamten Entropie darstellen. Die Minimierung der gemeinsamen Information als Ziel der blinden Quellentrennung läßt sich dementsprechend entweder durch die Minimierung der Entropien der einzelnen Kanäle oder durch die Maximierung der Entropie der multivariaten Verteilung erreichen.

4.2.2.1 Konzept der Minimierung der Einzelentropien

Unter der Voraussetzung, daß gaußverteilte Zufallsvariablen bei gegebener Varianz die größte Entropie aufweisen, läßt sich schlussfolgern, daß eine kleine Entropie gleichbedeutend mit einer Nichtgaußförmigkeit der Signale ist. Folglich führt eine möglichst große Abweichung von der Normalverteilung zu statistisch unabhängigen Zufallsvariablen. Dieses Konzept deckt sich mit der Aussage des zentralen Grenzwertsatzes, daß die Überlagerung unabhängiger Zufallsvariablen zu einer gaußförmigeren Verteilungsdichte führt. Die Minimierung der Entropie resultiert in einer weniger gaußförmigen Verteilungsdichte und weist damit auf Signale, die nicht aus der Überlagerung unabhängiger Quellsignale stammen.

Die Entropieminimierung eines einzelnen Kanals kann nach Gleichung (4.42) auch als Maximierung der Negentropie formuliert werden. Dabei kann die Negentropie in der Praxis durch Kumulanten approximiert werden, indem die Verteilungsdichtefunktion in der Nähe einer Gaußverteilung durch eine Gram-Charlier-Entwicklung mit (orthonormalen) Tschebyscheff-Hermitischen Polynomen dargestellt wird. Somit ergibt sich die Negentropie in Abhängigkeit vom Moment dritter Ordnung und der Kurtosis κ zu [47]

$$\mathcal{N}(x) \approx \frac{1}{12}E[x^3]^2 + \frac{1}{48}\kappa(x)^2. \quad (4.47)$$

Um die Sensitivität der Negentropie-Schätzung gegenüber statistischen Ausreißern zu verringern, können anstatt der Kumulanten andere nicht-quadratische Funktionen für die Approximation benutzt werden. Bei Verwendung einer ungeraden und einer geraden Funktion G_1 bzw. G_2 , lassen sich, analog zur Verwendung der Kumulanten, die Schiefe und Spitzigkeit der Verteilung bewerten. Die Negentropie läßt sich dann approximieren als

$$\mathcal{N}(x) \approx k_1 E^2[G_1(x)] + k_2 \left(E[G_2(x)] - E[G_2(x_{\text{Gauß}})] \right)^2, \quad (4.48)$$

wobei k_1 und k_2 positive Konstanten sind, $x_{\text{Gauß}}$ eine mittelwertfreie Variable mit gaußförmiger Verteilungsdichte und Einheitsvarianz darstellt [47] und x ebenfalls mittelwertfrei und auf Varianz eins normiert ist. Bei der Wahl von

$$G_1(x) = \frac{1}{a_1} \log \cosh a_1 x, \quad (4.49)$$

$$G_2(x) = -\exp(-x^2/2), \quad (4.50)$$

mit $1 \leq a_1 \leq 2$ erhält man eine robuste Schätzung der Negentropie [47].

Die Maximierung der Negentropie kann dann z.B. durch ein einfaches Gradientenverfahren erfolgen. Eine schnellere Konvergenz wird allerdings durch den Einsatz eines Fixed-Point-Algorithmus zur Optimierung erreicht. Dieses Verfahren ist als FastICA bekannt [48], [46] und wurde bereits auf komplexwertige Signale erweitert [12].

4.2.2.2 Konzept der Maximierung der Gesamtentropie

Aus Gleichung (4.46) läßt sich schlußfolgern, daß die Reduktion der gemeinsamen Information auch durch die Maximierung der Verbundentropie erreicht werden kann. Dieser Ansatz wird auch als Informationsmaximierung bezeichnet. Er wurde nach verschiedenen Vorarbeiten [63],[59] von Bell und Sejnowski [9] auf die blinde Quellentrennung angewandt.

Zur Verdeutlichung der Funktionsweise sei der einkanalige Fall anhand eines konnektionistischen Neurons (vgl. Abbildung 4.5) betrachtet. Die Eingangssignale \mathbf{x} werden als $u = \mathbf{w}^T \mathbf{x}$ gewichtet addiert und anschließend einkanalig durch eine nichtlineare Funktion $g(u)$ transformiert.

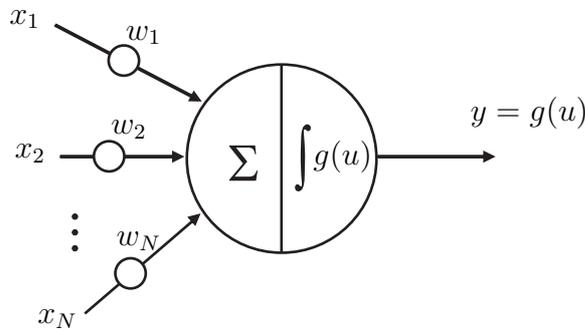


Abbildung 4.5: Konnektionistisches Neuron

Besitzt die Funktion $g(u)$ eine eindeutige Inverse, gilt für die Verteilungsdichten des Aus- und Eingangs der Zusammenhang [73]

$$p_y(y) = \frac{p_u(u)}{|\partial y / \partial u|} = \frac{p_u(u)}{|g'(u)|}. \quad (4.51)$$

Die Ausgangsentropie berechnet sich entsprechend Gleichung (4.41) als

$$H(y) = - \int_{-\infty}^{\infty} p_y(y) \log p_y(y) dy. \quad (4.52)$$

Durch Einsetzen von (4.51) in die obige Gleichung und mit $p_y(y)dy = p_u(u)du$ erhält man bei Annahme einer monoton steigenden Nichtlinearität

$$H(y) = - \int_{-\infty}^{\infty} p_u(u) \log \frac{p_u(u)}{g'(u)} du \quad (4.53)$$

$$= -D(p_u(u) || g'(u)) \leq 0. \quad (4.54)$$

Der Maximalwert der Entropie wird also genau dann erreicht, wenn die Ableitung der Funktion $g(u)$ mit der Verteilungsdichte des Signals übereinstimmt.

Im mehrkanaligen Fall wird der Eingangsvektor \mathbf{x} mit einer quadratischen Gewichtungs-
matrix \mathbf{W} multipliziert und Nichtlinearitäten werden auf alle Ausgangskanäle \mathbf{u} angewendet. Die transformierte Verteilungsdichte ist dann gegeben als

$$p_{\mathbf{y}}(\mathbf{y}) = \frac{p_{\mathbf{x}}(\mathbf{x})}{|\mathbf{J}|}, \quad (4.55)$$

wobei $|\mathbf{J}|$ die Determinante der Jacobimatrix

$$\mathbf{J} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_N} \\ \vdots & & \vdots \\ \frac{\partial y_N}{\partial x_1} & \cdots & \frac{\partial y_N}{\partial x_N} \end{bmatrix} \quad (4.56)$$

darstellt. Die Ausgangsentropie läßt sich damit als

$$H(\mathbf{y}) = -E[\log p_{\mathbf{y}}(\mathbf{y})] \quad (4.57)$$

$$= E[\log |\mathbf{J}|] - E[\log p_{\mathbf{x}}(\mathbf{x})] \quad (4.58)$$

formulieren. Für die Maximierung muß der Gradient bezüglich \mathbf{W} berechnet werden. Da nur der erste Term in (4.58) dazu beiträgt und die Maximierung mit einem stochastischen Gradientenverfahren erfolgt, bei dem der Erwartungswertoperator entfällt, ergibt sich für den Gradienten

$$\frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} = \frac{\partial}{\partial \mathbf{W}} \log |J|. \quad (4.59)$$

Mit

$$\det J = \det \mathbf{W} \prod_{i=1}^N |g'_i(u_i)| \quad (4.60)$$

folgt:

$$\frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} = \frac{\partial}{\partial \mathbf{W}} \log |\det W| + \frac{\partial}{\partial \mathbf{W}} \log \prod_i^N |y'_i|. \quad (4.61)$$

Nimmt man eine sigmoide Aktivierungsfunktion der Form $g(u) = (1 + e^{-u})^{-1}$ an, ergibt sich schließlich die Lernregel des InfoMax-Algorithmus [9]

$$\Delta \mathbf{W} \propto (\mathbf{W}^T)^{-1} + (\mathbf{1} - 2\mathbf{y})\mathbf{x}^T. \quad (4.62)$$

Durch die Anwendung des natürlichen bzw. relativen Gradienten [3],[25] kann eine schnellere Konvergenz erreicht werden. Die Lernregel verändert sich dann zu [50]

$$\Delta \mathbf{W} \propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W}. \quad (4.63)$$

4.2.3 ICA durch Maximum-Likelihood-Schätzung

Ein anderer weit verbreiteter Ansatz zur Bestimmung der unabhängigen Signalkomponenten ist durch die Anwendung der Maximum-Likelihood-Parameterschätzung gegeben. Ausgehend vom Mischungsmodell nach Gleichung (4.4), läßt sich die Transformation der Wahrscheinlichkeitsdichte der Quellsignale wie folgt darstellen [47]:

$$p_x(\mathbf{x}) = \frac{p_s(\mathbf{s})}{|\det \mathbf{A}|} = |\det \mathbf{W}| \cdot p_s(\mathbf{s}) = |\det \mathbf{W}| \prod_i p_i(s_i). \quad (4.64)$$

Dabei wurde vorausgesetzt, daß die Entmischungsmatrix $\mathbf{W} = \mathbf{A}^{-1}$ existiert, d.h. daß \mathbf{A} nicht singulär ist und daß die Quellsignale statistisch unabhängig voneinander sind, sich die Randdichteverteilungen also faktorisieren lassen.

Jedes einzelne Quellsignal läßt sich durch zeilenweise Multiplikation mit der Entmischungsmatrix aus den Sensorsignalen extrahieren. Mit \mathbf{w}_i als i -ter Zeile der Entmischungsmatrix $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_M]^T$ ergibt sich

$$p_x(\mathbf{x}) = |\det \mathbf{W}| \prod_i p_i(\mathbf{w}_i^T \mathbf{x}). \quad (4.65)$$

Die Likelihood-Funktion kann bei unabhängigen Beobachtungen als Produkt der Verteilungsdichten aufgestellt werden. Mit K als Anzahl der Beobachtungen folgt:

$$L(\mathbf{W}) = \prod_{k=1}^K \left[|\det \mathbf{W}| \prod_{i=1}^N p_i(\mathbf{w}_i^T \mathbf{x}(t)) \right]. \quad (4.66)$$

Aus rein praktischen Gesichtspunkten wird bevorzugt die Log-Likelihood-Funktion verwendet, da somit das Produkt der Verteilungen in eine Summe überführt werden kann

$$\log L(\mathbf{W}) = \sum_{k=1}^K \sum_{i=1}^N \log p_i(\mathbf{w}_i^T \mathbf{x}(t)) + K \log |\det \mathbf{W}|. \quad (4.67)$$

Üblicherweise wird die Summe über die Beobachtungen als Erwartungswert dargestellt und die Log-Likelihood-Funktion durch die Anzahl der Beobachtungen dividiert

$$\frac{1}{K} \log L(\mathbf{W}) = E \left[\sum_{i=1}^N \log p_i(\mathbf{w}_i^T \mathbf{x}(t)) \right] + \log |\det \mathbf{W}|. \quad (4.68)$$

Damit ist die Log-Likelihood-Funktion in Abhängigkeit von den Parametern als auch von den Verteilungsdichten der Quellen formuliert.

Allerdings ist die vollständige Schätzung der Verteilungsdichten auch in diesem Fall zu aufwendig für praktische Anwendungen. Aus diesem Grund müssen die Verteilungsdichten

durch Funktionen mit weniger Parametern approximiert werden. Es läßt sich zeigen [47], daß es dafür genügt, lediglich zwischen zwei Approximierenden zu unterscheiden, solange jeweils eine von beiden die Bedingung

$$E[s_i g_i(s_i) - g'_i(s_i)] > 0 \quad \forall i \quad (4.69)$$

erfüllt. Dabei stellt $g_i(s_i)$ die Ableitung der logarithmierten approximierenden Verteilungsdichte $\tilde{p}_i(s_i)$ dar

$$g_i(s_i) = \frac{\partial}{\partial s_i} \log \tilde{p}_i(s_i) = \frac{\tilde{p}'_i(s_i)}{\tilde{p}_i(s_i)}, \quad (4.70)$$

und die Ausgangssignale $y_i = \mathbf{w}_i^T \mathbf{x}$ müssen mit Prewhitening vorverarbeitet sein (vgl. Abschnitt 4.2.1.1). Bei Verwendung der Log-Verteilungsdichten

$$\log \tilde{p}_i^-(s_i) = \alpha_2 - (s_i^2/2 - \log \cosh(s_i)) \quad (4.71)$$

$$\log \tilde{p}_i^+(s_i) = \alpha_1 - 2 \log \cosh(s_i), \quad (4.72)$$

mit den positiven Parametern α_1 und α_2 , ist für eine der beiden sub- bzw. supergaußschen Verteilungen \tilde{p}^- und \tilde{p}^+ das Erfüllen der Anforderung (4.69) garantiert [47].

Die Maximierung der Likelihood-Funktion kann dann z.B. durch die Anwendung eines Gradientenverfahrens erfolgen. Der stochastische Gradient der Log-Likelihood-Funktion aus Gleichung (4.67) berechnet sich zu

$$\frac{1}{K} \frac{\partial \log L}{\partial \mathbf{W}} = (\mathbf{W}^T)^{-1} + E[\mathbf{g}(\mathbf{W}\mathbf{x})\mathbf{x}^T], \quad (4.73)$$

wobei die vektorielle Funktion $\mathbf{g}(\mathbf{y}) = [g_1(y_1), \dots, g_M(y_M)]$ die einzelnen Ableitungen der Log-Verteilungsdichten enthält

$$g_i(y_i) = (\log p_i)'. \quad (4.74)$$

Daraus folgt die Lernregel des Gradientenverfahrens:

$$\Delta \mathbf{W} \propto (\mathbf{W}^T)^{-1} + E[\mathbf{g}(\mathbf{W}\mathbf{x})\mathbf{x}^T]. \quad (4.75)$$

Bei Verwendung eines stochastischen Gradienten ergibt sich eine zum InfoMax-Algorithmus [9] identische Lernregel (vgl. Abschnitt 4.2.2.2), womit die enge Verwandtschaft zum Prinzip der Informationsmaximierung deutlich wird.

$$\Delta \mathbf{W} \propto (\mathbf{W}^T)^{-1} + \mathbf{g}(\mathbf{W}\mathbf{x})\mathbf{x}^T \quad (4.76)$$

Durch die Anwendung des natürlichen bzw. relativen Gradienten, der sich auch in diesem Fall durch Multiplikation mit $\mathbf{W}^T \mathbf{W}$ ergibt, läßt sich die Konvergenzgeschwindigkeit des Verfahrens erheblich steigern [47].

4.2.4 Kumulantenbasierte ICA-Verfahren

Eine weitere Methode der blinden Quellentrennung beruht auf der Auswertung der höheren Signalstatistik mit Hilfe klassischer statistischer Kenngrößen, wie Momente und Kumulanten, die aus der momentenerzeugenden Funktion bzw. der kumulantenerzeugenden Funktion abgeleitet werden können [33]. Von besonderer Bedeutung sind in diesem Zusammenhang Kumulanten vierter Ordnung. Dies liegt zum einen daran, daß Kumulanten noch höherer Ordnung unzuverlässiger zu schätzen sind und die meisten der in der Praxis auftretenden Signale eine symmetrische Verteilungsdichte aufweisen, deren Kumulante dritter Ordnung gleich Null ist. Zum anderen sind Kumulanten vierter Ordnung blind gegenüber additivem, gaußverteilten Rauschen [71].

Die allgemeine Beziehung zwischen Momenten und Kumulanten der Ordnung n lautet

$$\text{cum}(x_1, \dots, x_n) = \sum (-1)^{p-1} (p-1)! \cdot E\left[\prod_{i \in \mathfrak{s}_1} x_i\right] \cdot \dots \cdot E\left[\prod_{i \in \mathfrak{s}_p} x_i\right], \quad (4.77)$$

wobei die einzelnen \mathfrak{s}_i die Kombinationsmöglichkeiten bis hin zur Ordnung $p = n$ bezeichnen. Daraus folgt z.B. für die Kumulante vierter Ordnung die Vorschrift [71]:

$$\begin{aligned} \text{cum}(x_1, x_2, x_3, x_4) &= E[x_1 x_2 x_3 x_4] - E[x_1 x_2] \cdot E[x_3 x_4] - E[x_1 x_3] \cdot E[x_2 x_4] - E[x_1 x_4] \cdot E[x_2 x_3] \\ &\quad - E[x_1] \cdot [x_2 x_3 x_4] - E[x_2] \cdot [x_1 x_3 x_4] - E[x_3] \cdot E[x_1 x_2 x_4] - E[x_4] \cdot E[x_1 x_2 x_3] \\ &\quad + 2E[x_1 x_2] \cdot E[x_3] \cdot E[x_4] + 2E[x_1 x_3] \cdot E[x_2] \cdot E[x_4] + 2E[x_1 x_4] \cdot E[x_2] \cdot E[x_3] \\ &\quad + 2E[x_2 x_4] \cdot E[x_1] \cdot E[x_3] + 2E[x_3 x_4] \cdot E[x_1] \cdot E[x_2] + 2E[x_2 x_2] \cdot E[x_1] \cdot E[x_4] \\ &\quad - 6E[x_1] \cdot E[x_2] \cdot E[x_3] \cdot E[x_4]. \end{aligned} \quad (4.78)$$

Für mittelwertbefreite Signale, mit $E[x_i] = 0$, ergibt sich in vereinfachter Form

$$\begin{aligned} \text{cum}(x_1, x_2, x_3, x_4) &= E[x_1 x_2 x_3 x_4] - E[x_1 x_2] E[x_3 x_4] \\ &\quad - E[x_1 x_3] E[x_2 x_4] - E[x_1 x_4] E[x_2 x_3]. \end{aligned} \quad (4.79)$$

Gleichung (4.79) beschreibt einen vierdimensionalen Kumulantentensor, dessen Elemente an der Stelle (i, j, k, l) durch die entsprechende Kumulante gegeben sind. Der Tensor läßt sich auch durch einen Satz von Kumulantenmatrizen darstellen. Für eine $n \times n$ Matrix \mathbf{M} weist die Kumulantenmatrix $\mathbf{Q}(\mathbf{M})$ die Einträge

$$[\mathbf{Q}(\mathbf{M})]_{ij} = \sum_{k,l=1}^n m_{kl} \text{cum}(x_i, x_j, x_k, x_l) \quad \text{mit} \quad 1 \leq j, j \leq n \quad (4.80)$$

auf [23]. Der Vorteil dieser Darstellung liegt in der somit möglichen Formulierung der ICA-Aufgabenstellung als Eigenwertproblem.

Es läßt sich zeigen, daß nach der Vorverarbeitung der Sensorsignale die Kumulantenmatrix durch die orthogonale Entmischungsmatrix \mathbf{O} diagonalisiert wird [23]

$$\mathbf{O} \cdot \mathbf{Q}(\mathbf{M}) \cdot \mathbf{O}^T = \mathbf{\Lambda}_M \quad \text{mit} \quad \mathbf{\Lambda} = \text{diag}(\kappa_1 \mathbf{o}_1^T \mathbf{M} \mathbf{o}_1, \dots, \kappa_N \mathbf{o}_N^T \mathbf{M} \mathbf{o}_N). \quad (4.81)$$

Dieses Prinzip ist die Grundlage des *JADE*-Algorithmus (*Joint Approximate Diagonalization of Eigen-Matrices*), bei dem durch Auswahl der n signifikanten Eigenmatrizen (mit n als Anzahl der Quellen) nur ein beschränkter Satz an Kumulantenmatrizen diagonalisiert werden muß [24]. Diese Verbunddiagonalisierung läßt sich effektiv mit einem Jakobi-Verfahren [83] durchführen, bei dem in jedem Schritt die Bestimmung lediglich eines Rotationswinkels nötig ist.

Es läßt sich zeigen, daß die Verbunddiagonalisierung der Kumulantenmatrizen mit dem *JADE*-Algorithmus gleichbedeutend zur Minimierung der Summe der quadrierten Kreuzkumulanten der Ausgangssignale mit unterschiedlichem ersten und zweiten Index ist [47].

$$\mathcal{K}_{JADE} = \sum_{ijkl \neq iikl} |\text{cum}(y_i, y_j, y_k, y_l)|^2 \quad (4.82)$$

Dies korrespondiert mit der Tatsache, daß Kreuzkumulanten unabhängiger Zufallsvariablen stets gleich Null sind. Folglich stellt die Summe aller quadrierten Kreuzkumulanten ebenfalls ein Maß für die statistische Unabhängigkeit von Zufallsvariablen dar. Da sich die Summe aus quadrierten Kreuzkumulanten plus quadrierter Auto-Kumulanten bei Anwendung einer orthogonalen Transformation nicht ändert, kann die Minimierung der Summe aller quadrierter Kreuzkumulanten auch als Maximierung von

$$\mathcal{K}_{ICA} = \sum_{i=1}^n |\text{cum}(y_i, y_i, y_i, y_i)|^2 \quad (4.83)$$

formuliert werden [26]. Die asymptotische Qualität der beiden Ansätze ist nach [22] gleich.

Eine zu diesen Verfahren eng verwandte Methode, die *Fourth Order Blind Identification* (FOBI), basiert auf der Eigenwertzerlegung von

$$\mathbf{\Omega} = E[\mathbf{x}_s \mathbf{x}_s^T \|\mathbf{x}_s\|^2], \quad (4.84)$$

einer Korrelationsmatrix vierter Ordnung. Für den eingeschränkten Anwendungsfall von Quellen mit unterschiedlicher Kurtosis stellt diese Methode einen der effizientesten Ansätze überhaupt dar [47], da die orthogonale Mischungsmatrix direkt aus der Eigenwertzerlegung bestimmt werden kann.

Wie schon in Abschnitt 4.2.2.1 gezeigt wurde, lassen sich mit Hilfe von Kumulanten noch weitere Maße zur Bewertung der statistischen Unabhängigkeit approximieren [25]. Dabei wird der enge Zusammenhang zu den informationstheoretisch motivierten Verfahren und der Maximum-Likelihood-Quellentrennung aus Abschnitt 4.2.2 bzw. 4.2.3 deutlich.

4.2.5 Verfahren auf der Basis von Statistik zweiter Ordnung

Allen bisher beschriebenen Verfahren ist gemeinsam, daß sie zur Bewertung der statistischen Unabhängigkeit direkt oder indirekt statistische Kenngrößen höherer Ordnung ausnutzen. Für den Spezialfall gaußverteilter Zufallsvariablen, die sich vollständig mit statistischen Kenngrößen zweiter Ordnung (Mittelwert und Varianz) beschreiben lassen, sind die statistischen Kenngrößen höherer Ordnung gleich Null.

Da Überlagerungen gaußscher Variablen wieder gaußsche Variablen ergeben, können Mischungen dieser Art nicht rekonstruiert werden. Diese Einschränkung läßt sich nur durch die Hinzunahme weiterer Informationen umgehen. Dafür existieren Ansätze zur Ausnutzung der Nichtstationarität [66], des spektralen Gehaltes der Quellsignale [53] oder auch beider Kriterien zusammen [19].

Im Bereich der Sprachsignalverarbeitung ist insbesondere die Verwendung zeitlicher bzw. spektraler Signaleigenschaften weit verbreitet. Das zugrunde liegende Prinzip ist dabei, daß für unabhängige Signale die Kreuzkorrelation zu allen Zeitpunkten verschwindet

$$E[s_i(t)s_j(t + \tau)] = 0. \quad (4.85)$$

Gleichzeitig müssen die Signale eine spektrale Färbung aufweisen

$$E[s_i(t)s_i(t + \tau)] \neq 0. \quad (4.86)$$

Folglich sind die zu diesen Zeitpunkten bestimmten Kovarianzmatrizen diagonal

$$\mathbf{C}(t, \tau) = E[\mathbf{s}(t)\mathbf{s}^T(t + \tau)] = \mathbf{\Lambda}(t, \tau). \quad (4.87)$$

Mit dem Modell der linearen instantanen Überlagerung nach Gleichung (4.4) ergibt sich daher für die Sensorsignale

$$\mathbf{C}_x(t, \tau) = E[\mathbf{x}(t)\mathbf{x}^T(t + \tau)] = E[\mathbf{A}\mathbf{s}(t)\mathbf{A}\mathbf{s}^T(t + \tau)] = \mathbf{A}\mathbf{\Lambda}(t, \tau)\mathbf{A}^T. \quad (4.88)$$

Durch die Auswahl von genau zwei verschiedenen Kovarianzmatrizen zu den Zeitpunkten τ_i und τ_j läßt sich ein Eigenwertproblem der Form

$$(\mathbf{C}_i\mathbf{C}_j^{-1})\mathbf{A} = \mathbf{A}(\mathbf{\Lambda}_i\mathbf{\Lambda}_j^{-1}) \quad (4.89)$$

aufstellen, um die Mischungsmatrix zu berechnen [68]. Wegen der kritischen Wahl einzelner τ ist es jedoch günstiger, mehrere Zeitpunkte zu berücksichtigen. Nach der aus Abschnitt 4.2.1.1 bekannten Vorverarbeitung mit

$$\mathbf{x}_s = \mathbf{O}\mathbf{s}$$

reduziert sich die Suche auf den Raum orthogonaler Matrizen. Es muß eine orthogonale Matrix gefunden werden, die den Satz an Kovarianzmatrizen

$$\mathbf{C}_{x_s}(t, \tau) = E[\mathbf{x}_s(t)\mathbf{x}_s^T(t + \tau)] = \mathbf{O}\mathbf{\Lambda}(t, \tau)\mathbf{O}^T \quad (4.90)$$

für alle betrachteten Zeitpunkte wie folgt diagonalisiert [10]:

$$\mathbf{O}^T\mathbf{C}_{x_s}(t, \tau)\mathbf{O} = \mathbf{\Lambda}(t, \tau). \quad (4.91)$$

Diese Problemstellung ist bereits von der Verbunddiagonalisierung der Kumulantenmatrizen aus Abschnitt 4.2.4 bekannt. Folglich kann auch in diesem Fall ein Jakobi-Verfahren für die Diagonalisierung der Kovarianzmatrizen benutzt werden. Die Methode der Verbunddiagonalisierung eines Satzes von zeitverzögerten Kovarianzmatrizen wird auch als *Second order Blind Identification* (SOBI) bezeichnet [10].

Eine Erweiterung bzw. Verallgemeinerung des spektralen Konzepts stellt der Einsatz linearer Operatoren der Form

$$\mathbf{T}(\mathbf{s}) = \mathbf{T}([s_1, \dots, s_N]) = [T_1(s_1), \dots, T_N(s_N)] \quad (4.92)$$

dar [53]. Da sie die statistischen Abhängigkeiten zwischen den Sensorsignalen nicht verändern, aber bei einer geeigneten Wahl die speziellen spektralen Eigenschaften der Signale besser herausstellen, lassen sie sich vorteilhaft zur blinden Quellentrennung benutzen. In [53] werden dazu insbesondere die Wavelet-Transformation, Signalangepaßte Filter, optimale FIR-Filter sowie Autoregressive (AR) Modelle verwendet.

4.3 Blinde Quellentrennung konvolutiver Mischungen

Wie bereits in Abschnitt 4.1.2 deutlich wurde, stellen konvolute Mischungen eine Überlagerung gefilterter Quellsignale dar. Zur Trennung einer solchen Mischung ist es notwendig, eine im Vergleich zu instantanen Überlagerungen sehr viel größere Anzahl an Parametern zu schätzen, da für jedes der M Quellsignale mindestens $M - 1$ Mischungsfilter berechnet werden müssen und ein einzelnes Filter mehrere Tausend Koeffizienten aufweisen kann.

Um diesen Aufwand zu reduzieren und die Konvergenz der Verfahren zu beschleunigen, basieren die meisten Algorithmen zur Trennung konvolutiver Mischungen auf der in Abschnitt 4.1.2 beschriebenen Verarbeitung im Frequenzbereich. Abbildung 4.6 auf der nächsten Seite verdeutlicht schematisch die zur Quellentrennung notwendigen Schritte.

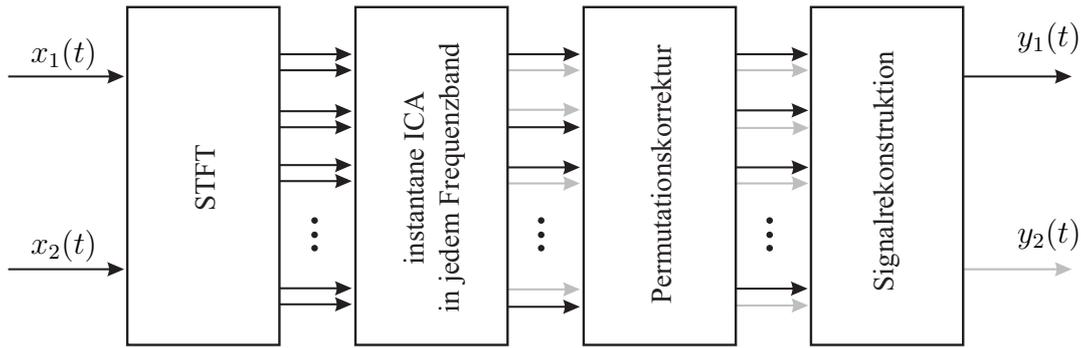


Abbildung 4.6: Funktionsblöcke für frequenzbereichsbasierte ICA-Verfahren

Nach der Kurzzeit-Fouriertransformation muß für jede Frequenz ein komplexwertiges, instantanes Mischungsproblem gelöst werden. Da die in jedem Frequenzband gefundenen Lösungen unabhängig voneinander sind, können einzelne Frequenzbänder vertauscht sein. Um daraus entstehende Fehler zu minimieren, wird anschließend eine Stufe zur Permutationskorrektur eingesetzt. Die Signalrekonstruktion erfolgt dann entweder durch Rücktransformation der Entmischungsfiler in den Zeitbereich und FIR-Filterung oder durch direkte Entmischung der Signale im Frequenzbereich und anschließende Rücktransformation in den Zeitbereich.

Die ersten Verfahren dieser Art entstanden in den 90-er Jahren (siehe z.B. [20], [34], [70], [60]). Sie waren jedoch für reale Mikrofonaufnahmen mit den z. T. sehr langen Mischungsfiltern nicht geeignet. Die Ursache dafür ist vor allem in der unzureichenden Permutationskorrektur zu sehen.

Eine wesentliche Verbesserung dazu stellte das Verfahren nach Parra und Spence dar [74]. Dieser Algorithmus wertet in jedem Frequenzband die statistische Unabhängigkeit nichtstationärer Signale, wie z.B. Sprache, aus, indem das Kreuzleistungsdichtespektrum für mehrere zeitlich versetzte Intervalle bewertet wird. Die mit Hilfe einer bestimmten Anzahl an Zeitpunkten in einem Intervall geschätzte Matrix der Kreuzleistungsdichtespektren der Sensorsignale $\hat{\Phi}_{xx}(\omega, t)$ muß bei unabhängigen Quellsignalen und additivem Rauschen dem Modell

$$\hat{\Phi}_{xx}(\omega, t) = \mathbf{A}(\omega)\mathbf{\Lambda}_s(\omega, t)\mathbf{A}^H(\omega) + \mathbf{\Lambda}_n(\omega, t) \quad (4.93)$$

genügen, wobei die diagonalen Matrizen $\mathbf{\Lambda}_s(\omega, t)$ und $\mathbf{\Lambda}_n(\omega, t)$ die Autoleistungsdichtespektren der Quellsignale bzw. des Rauschens darstellen. In Abhängigkeit von der Entmischungsmatrix \mathbf{W} läßt sich diese Beziehung auch als

$$\mathbf{\Lambda}_s(\omega, t) = \mathbf{W}(\omega)(\hat{\Phi}_{xx}(\omega, t) - \mathbf{\Lambda}_n(\omega, t))\mathbf{W}^H(\omega) \quad (4.94)$$

formulieren. Damit kann in jedem Frequenzband eine Fehlerfunktion aufgestellt werden, welche die Abweichung von diesem Modell beschreibt

$$\varepsilon(\omega, t) = \mathbf{W}(\omega)(\hat{\Phi}_{xx}(\omega, t) - \Lambda_n(\omega, t))\mathbf{W}^H(\omega) - \Lambda_s(\omega, t). \quad (4.95)$$

Auf dieser Grundlage wertet die Kostenfunktion das Betragsquadrat der Fehlerfunktion für alle Frequenzen und die insgesamt K Intervalle aus

$$\mathcal{K} = \sum_{\omega=1}^{L/2} \sum_{k=1}^K \|\varepsilon(\omega, k)\|^2. \quad (4.96)$$

Die Optimierung der unbekanntenen Parameter \mathbf{W} , Λ_s und Λ_n erfolgt schließlich mittels Gradientenabstiegsverfahren.

Die wesentliche Neuerung dieses Verfahrens besteht jedoch in der in jedem Zwischenschritt durchzuführenden zeitlichen Begrenzung der Entmischungsfiler. D.h. die durch Entmischungsmatrizen gebildeten Übertragungsfunktionen werden in den Zeitbereich transformiert, stückweise auf Null gesetzt und anschließend wieder in den Frequenzbereich transformiert. Die zugrunde liegende Annahme ist dabei, daß zeitlich kurze Entmischungsfiler einen glatten Verlauf im Frequenzbereich aufweisen, der bei eventuell auftretenden Permutationen nicht zu erwarten ist. Die Erzwingung kurzer Filterlängen hat sich als sehr wirksam herausgestellt und wurde deshalb in einigen anderen Verfahren übernommen (siehe z.B. [30], [99]).

Die Forderung nach glatten Filtern kann auch direkt im Frequenzbereich zur Korrektur der Permutationen genutzt werden. Dabei erfolgt die Zuordnung anhand eines Ähnlichkeitskriteriums benachbarter Frequenzen [20], [8].

Der Vollständigkeit halber sei angemerkt, daß es auch Verfahren zur Trennung konvolutiver Mischungen gibt, die aufgrund spezieller Kostenfunktionen ohne Permutationskorrektur auskommen. In [4] wird die über mehrere Frequenzbänder gleiche Amplitudenmodulation von Sprachsignalen zur Entmischung benutzt. Da alle Frequenzbänder gleichzeitig betrachtet werden, ist keine Notwendigkeit zur Permutationskorrektur gegeben. Der Rechenaufwand für dieses Verfahren ist jedoch vergleichsweise hoch.

Ein vollständig anderer Ansatz ist in [78] beschrieben. Dabei wird angenommen, daß zu einem bestimmten Zeitpunkt und für eine Frequenz nur eines der Quellsignale im Sensorsignal vorkommt. Durch die Konstruktion einer geeigneten Zeit-Frequenz-Maske lassen sich deshalb die Quellsignale einzeln rekonstruieren. Der Vorteil dieses Verfahrens ist, daß sich damit bei geringem Rechenaufwand auch unterbestimmte Mischungsprobleme lösen lassen.

4.3.1 Kombination von ICA und Beamforming

Neben den im vorigen Abschnitt vorgestellten Kriterien der Glattheit bzw. der kurzen Entmischungsfiler lassen sich aufgrund der mehrkanaligen Signalerfassung auch räumliche Informationen für die Verbesserung der Trennungsleistung von Algorithmen zur konvolutiven blinden Quellentrennung ausnutzen.

So beschreibt [8] ein Multi-DSP-System zur blinden Quellentrennung mit vorgeschaltetem Delay-And-Sum-Beamformer, der für eine Glättung der Übertragungsfunktion in Blickrichtung sorgen soll. In [52] wird der Zusammenhang zwischen Minimum-Variance-Beamforming und PCA untersucht und darauf basierend ein BSS-Verfahren mit erhöhter Robustheit vorgestellt. Der Algorithmus nach [80] wechselt im jeweiligen Frequenzband iterativ zwischen ICA und Nullbeamforming und erreicht damit bessere Trennungsergebnisse als mit herkömmlichen Verfahren der blinden Quellentrennung.

Ein kombinierter Ansatz wird in [75] vorgeschlagen: In den im vorigen Kapitel beschriebenen Algorithmus nach Parra [74] werden verschiedene, räumliche Nebenbedingungen integriert. Zum einen wird die Einhaltung einer ungestörten Übertragungsfunktion in Blickrichtung und zum anderen zusätzlich das Ausnullen der jeweils anderen Blickrichtungen gefordert. Obwohl die modellhafte Annahme konstanter Einfallrichtungen in der Realität nicht zutrifft, können auch hier verbesserte Trennungsergebnisse erzielt werden.

Zusammenfassend läßt sich sagen, daß die erreichten Verbesserungen dieser Verfahren im allgemeinen nicht auf die bessere Adaption der Filterkoeffizienten zurückzuführen sind. Vielmehr ist durch die Einbeziehung räumlicher Informationen ein effektives Kriterium zur Lösung des Permutationsproblems gegeben. Unter der Voraussetzung kleiner Mikrofonabstände und bei einer begrenzten Anzahl an Quellen stellt diese Methode derzeit den zuverlässigsten Ansatz zur Lösung des Permutationsproblems dar.

4.3.1.1 Beampattern für Permutationskorrektur

Aus den vorherigen Betrachtungen wird deutlich, daß sich die Trennungsqualität der blinden Quellentrennung gefalteter Mischungen insbesondere dann erhöhen läßt, wenn räumliche Informationen für die Lösung des Permutationsproblems ausgenutzt werden. Eine Möglichkeit dafür ist, die durch ICA-Methoden gefundenen Entmischungsfiler direkt als Beampattern auszuwerten und die Frequenzbänder anschließend anhand übereinstimmender Richtungen zu sortieren.

Die ersten Veröffentlichungen zu diesem Thema stammen von Kurita [55, 56]. Es wird vorgeschlagen, mit Hilfe des Beampatterns sowohl das Permutations- als auch das Skalierungsproblem eines Frequenzbereich-basierten ICA Algorithmus zu lösen. Die Zuordnung der Quellen orientiert sich dabei an den Ausnullungsrichtungen, wodurch in verhalten Umgebungen eine im Vergleich zu anderen Verfahren verbesserte Störgeräuschunterdrückung erreicht wird.

Für die Berechnung der Beampattern muß der frequenz- und richtungsabhängige Phasenvektor, definiert als $\mathbf{r}(\omega, \varphi) = [1, \dots, r_M(\omega, \phi)]^T$, mit den durch die Zeilen der Entmischungsmatrix $\mathbf{W}(\omega)$ gebildeten Filtern multipliziert werden. Für den 2×2 -Fall erhält man die zwei Beampattern

$$\Psi_1(\omega, \phi) = [W_{11}(\omega) \ W_{12}(\omega)] \cdot \mathbf{r}(\omega, \phi) \quad (4.97)$$

$$\Psi_2(\omega, \phi) = [W_{21}(\omega) \ W_{22}(\omega)] \cdot \mathbf{r}(\omega, \phi). \quad (4.98)$$

Zur Veranschaulichung sind in Abbildung 4.7 beispielhaft die Beträge der zwei Beampattern von unsortierten ICA-Entmischungsfiltern dargestellt. Dabei wurden in diesem Fall ausschließlich die Phasenanteile von $\mathbf{W}(\omega)$ berücksichtigt, um die Ausbildung der räumlichen Nullen besser zu verdeutlichen.

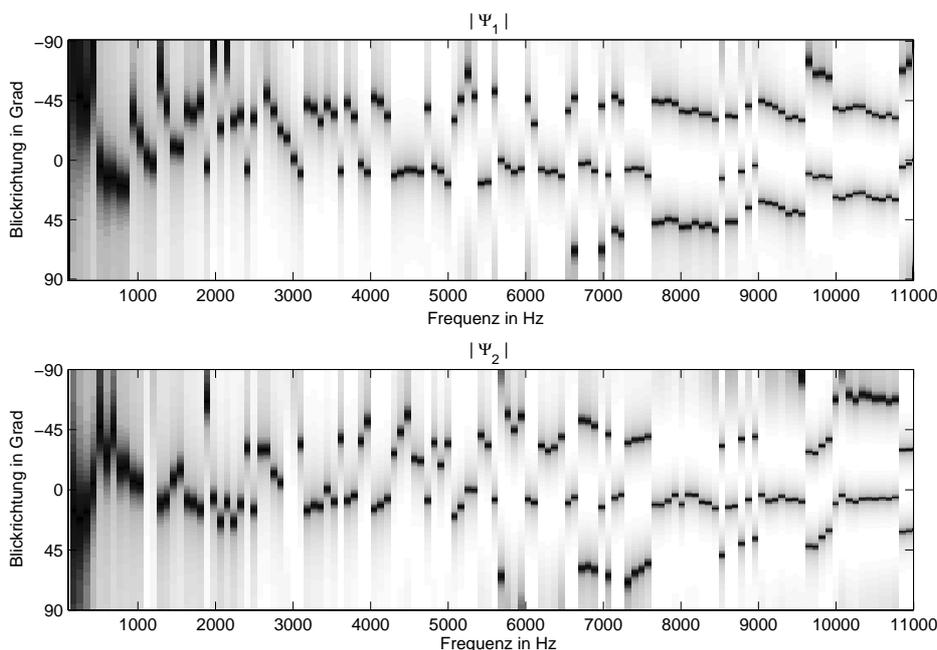


Abbildung 4.7: Beispiel für Beampattern von ICA-Entmischungsfiltern bei einem Mikrofonabstand von $d = 3$ cm und Quellensignalen aus 20° und -40°

Die Gegenüberstellung der Richtungen mit der größten Dämpfung, die sich zu

$$\phi_{0_i}(\omega) = \arg \min \Psi_i(\omega, \phi) \quad (4.99)$$

berechnen, zeigt Abbildung 4.8. Es sind prinzipiell die beiden Einfallsrichtungen der Quellsignale bei ca. 20° und -40° zu erkennen, wobei deutlich wird, daß die Richtungen der Nullen in einigen Frequenzbändern vertauscht sind. Weiterhin ist zu sehen, daß aufgrund des gewählten Mikrofonabstandes von 3 cm ab etwa 5.6 kHz räumliches Aliasing auftritt, wodurch sich die Position der nach Gleichung (4.99) berechneten Nullstellen in den oberen Frequenzbändern stark verändern kann.

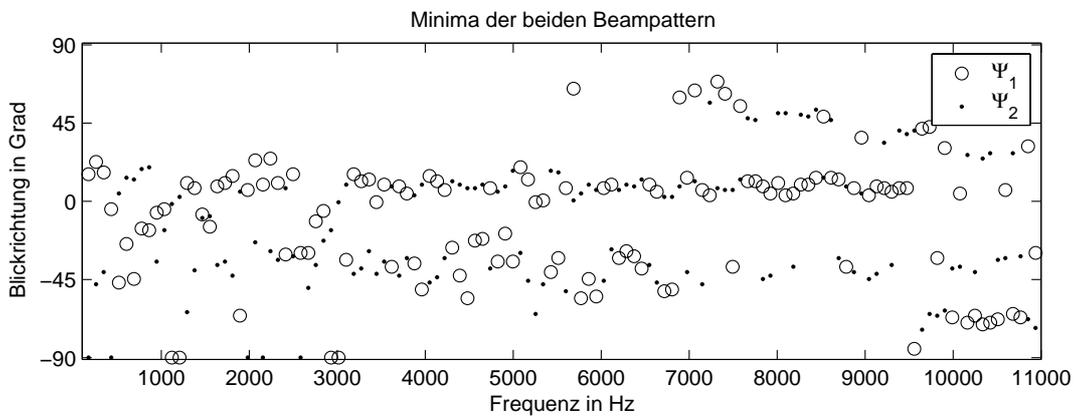


Abbildung 4.8: Gegenüberstellung der Minima der beiden Beampattern aus Abbildung 4.7

Um die daraus möglicherweise entstehenden Zuordnungsfehler zu minimieren, wird in [7] eine robustere Permutationskorrektur sowohl unter Berücksichtigung der zu beobachtenden Streuung der Richtungen über der Frequenz als auch unter Berücksichtigung möglicher Alias-Winkel vorgeschlagen. Dabei wird angenommen, daß das mit Hilfe von Gleichung (4.99) berechnete Minimum nicht zwangsläufig die korrekte Einfallsrichtung anzeigt, sondern bei Verletzung der Bedingung aus Gleichung (3.20) aus einem Satz möglicher Minima stammt, deren Lage durch die Parameter der Unterabtastung der sich ausbreitenden Welle bestimmt ist.

Aufgrund der 2π -Periodizität der komplexen Exponentialfunktion, die den Phasenversatz zwischen den Sensoren beschreibt, lautet der Zusammenhang zwischen beobachteter Einfallsrichtung ϕ_0 und möglicher Alias-Einfallsrichtung ϕ_{alias}

$$k \cdot 2\pi + \exp\left(-i2\pi \frac{d}{\lambda} \sin \phi_0\right) = \exp\left(-i2\pi \frac{d}{\lambda} \sin \phi_{alias}\right) \quad \text{mit } k \in \mathbb{Z}. \quad (4.100)$$

Durch Umformen läßt sich der Satz möglicher Einfallrichtungen angeben als

$$\phi_{alias} = \text{asin}\left(\frac{k \cdot \lambda}{d} + \sin \phi_0\right) \quad \text{mit } k \in \mathbb{Z}. \quad (4.101)$$

Unter Einbeziehung der auf diese Weise berechneten Einfallrichtungen kann eine Maximum-Likelihood-Zuordnung der Minima aller Frequenzbänder erfolgen, für deren Streuung eine Gaußverteilung zugrunde gelegt ist. Die Anpassung der Mittelwerte und Varianzen erfolgt iterativ mit Hilfe des EM-Algorithmus (*Expectation Maximization*). Unzuverlässige Minima, die sich nicht eindeutig zuordnen lassen, werden durch ihre gemittelten benachbarten Werte ersetzt. Liegt der Mikrofonabstand in der Größenordnung der Wellenlänge der höchsten zu verarbeitenden Frequenz, führt dieser Ansatz innerhalb weniger Iterationen zu einer robusten Zuordnung der Richtungen, wie aus Abbildung 4.9 deutlich wird.

Abschließend sei bemerkt, daß sich der Bereich möglicher Einfallrichtungen in Abhängigkeit von den akustischen Bedingungen verbreitern kann (siehe Abschnitt 5.2.5.2). Für Räume mit sehr langen Nachhallzeiten wird die richtungsbasierte Permutationskorrektur weniger zuverlässig, und es müssen zusätzliche Kriterien für das Sortieren der Entmischungsfiler herangezogen werden.

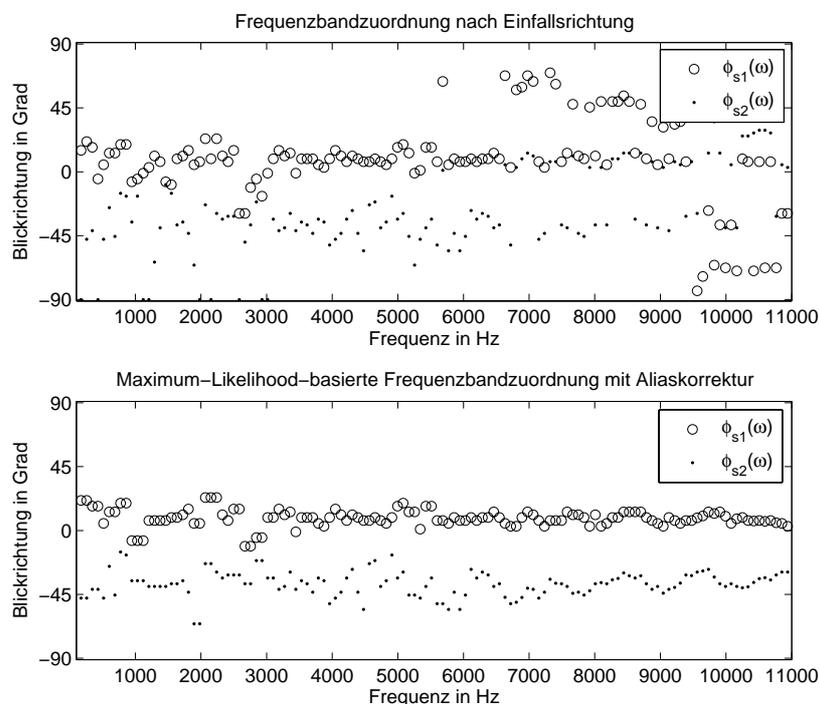


Abbildung 4.9: Gegenüberstellung der Permutationskorrektur durch Auswertung der Minima der Beampattern (oben) sowie durch Maximum-Likelihood-Zuordnung nach [7] (unten)

5 Optimierung frequenzvarianter Nullbeamformer mittels Statistik höherer Ordnung

5.1 Vergleich von Beamforming und konvolutiver ICA

Für die Entwicklung eines robusten Verfahrens zur Mehrkanal-Störgeräuschunterdrückung, das die Vorteile von Beamformern und den Methoden der blinden Quellentrennung in sich vereint, ist es zunächst notwendig, beide Ansätze miteinander zu vergleichen. Dabei ist eine klare Differenzierung jedoch nicht immer möglich. Prinzipiell könnten alle Ansätze der mehrkanaligen Störgeräuschunterdrückung als Beamforming bezeichnet werden, da sich auch BSS-Verfahren als räumliche Filter interpretieren lassen. Die signifikanten Unterschiede zwischen beiden Verfahren liegen vor allem in der Anzahl der Parameter, in der Struktur sowie in der Kostenfunktion.

5.1.1 Parametervergleich

Bei der Gegenüberstellung der Verfahren bezüglich der Anzahl der freien Parameter wird nachfolgend neben der blinden Quellentrennung auch zwischen adaptiven und nicht adaptiven Beamforming-Methoden (siehe Kapitel 3) unterschieden.

Fest entworfene Beamformer arbeiten signalunabhängig, d.h. für die Berechnung des Ausgangssignals sind nur die im Vorfeld getroffenen Annahmen der räumlichen Quellenverteilung von Bedeutung. Sofern diese Annahmen zutreffend sind, könnte mit nicht adaptiven Verfahren die gleiche Störsignaldämpfung wie mit allen anderen betrachteten Verfahren erreicht werden. Davon kann man in der Praxis aber nicht ausgehen, da insbesondere aufgrund der in realen Räumen auftretenden Nachhallzeiten weder eine absolute räumliche Zuordnung der Quellen noch die Festlegung der zu erwartenden Dämpfung möglich ist. Das hat zur Konsequenz, daß man für eine hohe Störsignaldämpfung ganze

Tabelle 5.1: Anzahl der freien Parameter von Beamforming und blinder Quellentrennung

| Verfahren | zu adaptierende Parameter |
|--------------------------|---|
| Nichtadaptive Beamformer | keine freien Parameter |
| Adaptive Beamformer | Vorgabe der Blickrichtung; alle anderen Modellparameter müssen gelernt werden |
| Blinde Quellentrennung | keine Vorgabe einer Modellgröße; alle Parameter müssen gelernt werden |

räumliche Bereiche (vgl. Abschnitt 3.3.2.2) ausblenden muß. Weiterhin müssen die verwendeten Mikrofone und Verstärker sehr genau sein, da auch diese Einflüsse a priori nicht bekannt sind.

Will man sämtliche mögliche Konfigurationen von Quellsignalpositionen, Nachhallzeiten, Verstärkungsfaktoren etc. berücksichtigen, erhöht sich die Anzahl der notwendigen Sensoren, insbesondere bei breitbandigen Schallsignalen, beträchtlich. Für eine zweikanalige Störgeräuschunterdrückung sind fest entworfene Beamformer deshalb ungeeignet.

Adaptive Beamformer passen bei vorgegebener Einfallsrichtung des Nutzsignals die Filterkoeffizienten an die Raumbedingungen an. Dadurch kommt man bei M Störsignalen mit $M + 1$ Mikrofonen aus. Im zweikanaligen Fall, bei genau einem kohärenten Quell- und Störsignal, müßte pro Frequenzband nur ein komplexer Parameter gelernt werden.

Allerdings ist auch bei adaptiven Beamformern die feste Vorgabe der Einfallsrichtung des Nutzsignals problematisch, sobald die Raumbedingungen vom idealen Fernfeld abweichen. Da mit diesem Problem in realen Umgebungen immer zu rechnen ist, kommt es bei adaptiven Beamformern oft zu einer nachhaltigen Störung des Nutzsignals.

Die blinde Quellentrennung benötigt prinzipiell die gleiche Anzahl an Sensoren wie adaptive Beamformer und muß aufgrund der im allgemeinen fehlenden räumlichen Nebenbedingung alle notwendigen Parameter adaptieren. Für den zweikanaligen Fall wäre pro Frequenzband die Erlernung von zwei komplexen Parametern nötig. Wie bei allen adaptiven Verfahren erhöht sich mit der Anzahl der freien Parameter der Rechenaufwand und u.U. auch die Adaptionzeit.

In der fehlenden Vorgabe der Nutzsignalrichtung liegt gleichzeitig die Stärke und ein Nachteil der blinden Quellentrennung. Der Vorteil ist, daß bei erfolgreicher Adaption nur eine äußerst geringe Nutzsignaldämpfung auftritt. Nachteilig ist die bei der frequenzbandweisen Berechnung fehlende räumliche Zuordnung der Quellsignale, die zum Permutationsproblem führt.

5.1.2 Strukturvergleich

Der wesentliche strukturelle Unterschied zwischen adaptiven Beamformern und Verfahren der blinden Quellentrennung für konvolute Mischungen besteht offensichtlich in der Anzahl der Ausgangssignale. Während Beamformer auf die Verbesserung eines bestimmten Quellsignals abzielen, ist man bei der blinden Quellentrennung eher an der Rekonstruktion aller an der Mischung beteiligten Signalkomponenten interessiert. In diesem Sinne läßt sich eine Struktur zur Trennung gefalteter Mischung auch als eine Parallelschaltung von zwei Filter-&-Sum-Beamformern auffassen, wie die nachfolgende Abbildung 5.1 für den zweikanaligen Fall verdeutlichen soll.

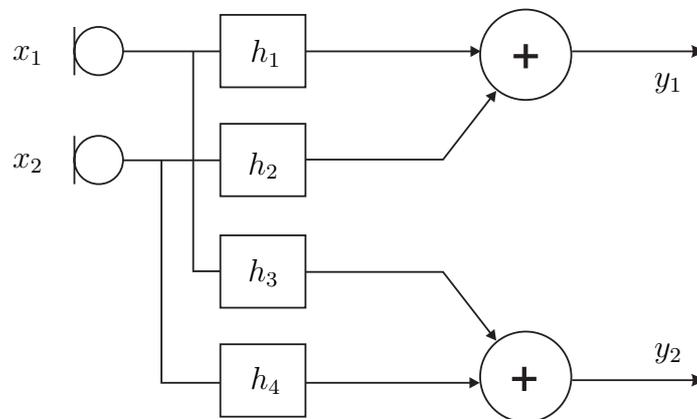


Abbildung 5.1: Konvolute Quellentrennung als Parallelschaltung von zwei Filter-And-Sum-Beamformern

Die beiden Ausgangssignale ergeben sich dabei zu

$$y_1(t) = h_1 * x_1(t) + h_2 * x_2(t)$$

$$y_2(t) = h_3 * x_1(t) + h_4 * x_2(t)$$

und entsprechen damit einer Feed-Forward-Struktur der konvolutiven blinden Quellentrennung.

5.1.3 Vergleich der Kostenfunktionen

Typischerweise wird bei adaptiven Beamformern eher Statistik zweiter Ordnung zur Parameteradaptation benutzt und z.B. die Ausgangsleistung unter Einhaltung diverser Nebenbedingungen minimiert. Bei der Analyse der statistischen Unabhängigkeit als Grundlage der blinden Quellentrennung ist Statistik höherer Ordnung bzw. die Hinzunahme weiterer

Kriterien zwingend notwendig (siehe Kapitel 4). Dadurch wird eine bessere Anpassung an die tatsächlichen Parameter möglich, sofern die Schätzung der Signalstatistik zuverlässig erfolgt. Letzteres ist jedoch tendenziell mit einer im Vergleich zur Bestimmung von Statistik zweiter Ordnung höheren Anzahl an notwendigen Samples verbunden.

Zusammenfassend kann man ICA-basierte Verfahren der blinden Quellentrennung also als parallelgeschaltete Filter-&-Sum-Beamformer bezeichnen, die ohne Vorgabe räumlicher Randbedingungen sämtliche Koeffizienten nach dem Kriterium statistischer Unabhängigkeit im Verbund optimieren.

5.2 Frequenzvarianter Nullbeamformer mit ICA-Kostenfunktion

5.2.1 Motivation

Die blinde Quellentrennung hat als Verfahren das größte Potenzial, eine sehr gute Störsignaldämpfung ohne Beeinträchtigung des Nutzsignals zu erreichen. Hauptsächlich aufgrund des Permutationsproblems ist die tatsächlich erreichte Trennungsqualität jedoch längst nicht so gut wie theoretisch möglich. Aus diesem Grund muß ein robustes Verfahren zur Unterdrückung gerichteter Störer zum einen in der Lage sein, alle Modellparameter zu erlernen und zum anderen eine eindeutige Signalzuordnung zu ermöglichen.

Die logische Konsequenz aus diesen Anforderungen ist, in einem Verfahren zur blinden Quellentrennung räumliche Informationen, z.B. gegeben durch die Parametrisierung des Fernfeld-Beamformingmodells, mit einzubeziehen. Damit läßt sich nicht nur der Suchraum einschränken, sondern es wird auch eine eindeutige Signalzuordnung möglich.

Die Vorgabe fester Blickrichtungen für Nutz- oder Störsignal in Form von Laufzeitunterschieden ist aufgrund der unzureichenden Anpassung an die realen Umgebungsbedingungen nicht für eine optimale Störgeräuschunterdrückung geeignet. Diese Tatsache kann anhand des konvolutiven Mischungsmodells aus Abschnitt 4.1.2 auf Seite 45 verdeutlicht werden. Demnach überlagern sich die mit der jeweiligen Raumimpulsantwort gefalteten Signale $s_1(t)$ und $s_2(t)$ an beiden Mikrofonen. Das bedeutet, daß die Mikrofonensignale nicht als Summe von ausschließlich zeitverzögerten Quellensignalen angenommen werden können, da die Raumimpulsantworten nicht aus einem einzelnen Impuls, entsprechend der Laufzeit zwischen Quelle und Mikrofon, bestehen.

Die gleiche Betrachtung im Frequenzbereich ergibt, daß die Mikrofonsignale in jedem Frequenzband aus mit dem jeweiligen Koeffizienten der Übertragungsfunktion multiplizierten, überlagerten Quellsignalen bestehen. Im Fall der reinen Zeitverzögerung wären diese Koeffizienten Phasenfaktoren der Form $\exp(-i\omega\tau_d)$ mit festem Zeitversatz τ_d .

Die Annahme eines festen Zeitversatzes ist die Grundlage für die bei adaptiven Beamformern verwendete Nebenbedingung, indem für alle Frequenzen eine konstante Einfallsrichtung des Nutzsignals vorausgesetzt wird. Aus diesem Grund kann es bei adaptiven Beamformern zur bereits in Abschnitt 5.1.1 angesprochenen Nutzsignalstörung kommen.

Die vorangegangene Betrachtung macht die Notwendigkeit einer frequenzabhängigen Richtungssuche deutlich, die Grundlage des frequenzvarianten Nullbeamforming ist. Durch das Ansetzen einer frequenzvariablen Einfallsrichtung wird der Phasenanteil der Raumimpulsantworten, nicht jedoch mögliche Unterschiede im Betrag kompensiert. Für die Auslöschung eines Signals ist jedoch in erster Linie die exakte Laufzeitkompensation, also der exakte Phasenfaktor wichtig. Der Betrag spielt dabei eine untergeordnete Rolle, er kann aber in einfacher Weise als Erweiterung in das Verfahren mit einbezogen werden (siehe Abschnitt 5.2.5.4).

5.2.2 Funktionsprinzip

Grundlage des ICA-basierten frequenzvarianten Nullbeamforming ist das konvolutive Mischungsmodell aus Abschnitt 4.1.2, bei dem für jedes Frequenzband ein komplexwertiges, instantanes Mischungsproblem zu lösen ist. Um räumliche Informationen einzuarbeiten, müssen die Phasenanteile der Filterkoeffizienten, basierend auf der Parametrisierung des Fernfeld-Modells, als Blickrichtungen interpretiert werden.

Die zu den Blickrichtungen äquivalenten Laufzeitunterschiede sind, bezogen auf ein bestimmtes Referenzmikrofon, im Frequenzbereich wie folgt definiert:

$$e(\omega, \varphi) = \exp\left(-i\omega\frac{d}{c}\sin(\varphi(\omega))\right), \quad (5.1)$$

wobei $\varphi(\omega)$ die durch die Raumübertragungsfunktion veränderte Einfallsrichtung darstellt. Mit Hilfe dieser blickrichtungsabhängigen Phasenfaktoren können die an den Mikrofonen ankommenden Quellsignale beschrieben werden. Für den zweikanaligen Fall ergeben sich die Meßsignale in jedem Frequenzband zu

$$X_1(\omega, \tilde{t}) = S_1(\omega, \tilde{t}) + S_2(\omega, \tilde{t}) \quad (5.2)$$

$$X_2(\omega, \tilde{t}) = e(\omega, \varphi_1) \cdot S_1(\omega, \tilde{t}) + e(\omega, \varphi_2) \cdot S_2(\omega, \tilde{t}). \quad (5.3)$$

Dabei dient Mikrofon eins als Referenzpunkt mit einer definierten Zeitverzögerung von Null, d.h. beide Quellsignale werden dort als unverzögert angenommen. In Bezug dazu weist an Mikrofon zwei das Quellsignal $S_1(\omega, \tilde{t})$ einen komplexen Phasenversatz von $e_1(\omega, \varphi_1)$ und das Quellsignal $S_2(\omega, \tilde{t})$ einen komplexen Phasenversatz von $e_2(\omega, \varphi_2)$ auf.

In Matrixschreibweise kann der in dieser Weise definierte Mischungsprozeß als

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ e(\varphi_1) & e(\varphi_2) \end{bmatrix} \cdot \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} \quad (5.4)$$

bzw. als

$$\mathbf{X} = \mathbf{A}_{ph} \cdot \mathbf{S} \quad (5.5)$$

geschrieben werden, wobei die aus der diskreten Kurzzeit-Fouriertransformation resultierenden Indizes \tilde{t} der Übersichtlichkeit halber nicht weiter mitgeführt werden.

Die Umkehrung dieses Prozesses ist, wie bereits in Kapitel 4 beschrieben, durch die Bildung der Inversen der Mischungsmatrix, nachfolgend als \mathbf{W} bezeichnet, gegeben

$$\mathbf{W} = \begin{bmatrix} 1 & 1 \\ e(\varphi_1) & e(\varphi_2) \end{bmatrix}^{-1} = \frac{1}{e(\varphi_1) - e(\varphi_2)} \cdot \begin{bmatrix} -e(\varphi_2) & 1 \\ e(\varphi_1) & -1 \end{bmatrix}. \quad (5.6)$$

Die Ausgangssignale \mathbf{Y} berechnen sich damit in jedem Frequenzband zu

$$\mathbf{Y} = \mathbf{W} \cdot \mathbf{X}. \quad (5.7)$$

Mit

$$\mathbf{W} \cdot \mathbf{A}_{ph} = \mathbf{I}, \quad (5.8)$$

wobei \mathbf{I} die Einheitsmatrix darstellt, können die folgenden Anforderungen an die Entmischungsfiler abgeleitet werden:

$$1 = W_{11} + e(\varphi_1)W_{12} \quad (5.9)$$

$$0 = W_{11} + e(\varphi_2)W_{12} \quad (5.10)$$

und

$$0 = W_{21} + e(\varphi_1)W_{22} \quad (5.11)$$

$$1 = W_{21} + e(\varphi_2)W_{22}. \quad (5.12)$$

Dabei stellen die Gleichungen (5.9) und (5.10) die Vorschrift für einen Filter-&-Sum-Beamformer dar, der das Signal aus Blickrichtung φ_1 ungestört durchläßt und das Signal

aus Blickrichtung φ_2 auslöscht. Im Gegensatz dazu, läßt der durch den zweiten Satz an Gleichungen, (5.11) und (5.12), gegebene Beamformer das Signal aus der Einfallsrichtung φ_2 ungefiltert durch und löscht das Signal aus Blickrichtung φ_1 . Die beiden parallelgeschalteten Beamformer arbeiten also nicht unabhängig voneinander. Vielmehr ist die Nutzsignalrichtung des einen die Störsignalrichtung des anderen und vice versa.

Mit Gleichung (5.6) ist ein Entmischungsmodell gegeben, das im zugrundeliegenden 2×2 Fall von den zwei komplexen Parametern

$$e_1 \equiv e(\omega, \varphi_1) = \exp\left(-i\omega \frac{d}{c} \sin(\varphi_1(\omega))\right) \quad (5.13)$$

und

$$e_2 \equiv e(\omega, \varphi_2) = \exp\left(-i\omega \frac{d}{c} \sin(\varphi_2(\omega))\right) \quad (5.14)$$

abhängt, die in jedem Frequenzband auf die zwei reellwertigen Parameter $\varphi_1(\omega)$ und $\varphi_2(\omega)$ zurückgeführt werden können. Dabei ist zu beachten, daß, im Unterschied zu konventionellen Delay-&-Sum-Beamformern, die Blickrichtungen $\varphi_1(\omega)$ und $\varphi_2(\omega)$ nicht für alle Frequenzen als konstant angesetzt werden. Statt dessen müssen beide Parameter in jedem Frequenzband so bestimmt werden, daß die sich ergebenden Ausgangssignale statistisch unabhängig voneinander sind.

5.2.3 Kostenfunktion

Als Maß für die statistische Unabhängigkeit können verschiedene, bereits in Kapitel 4 besprochene Kriterien herangezogen werden. Hierbei eignen sich insbesondere Kumulanten [84],[94], da sie im Gegensatz zu anderen Methoden unabhängiger von der Quellsignalstatistik sind [15], [76] und vor allem, da Kumulanten höherer Ordnung ($N > 2$) die Fähigkeit besitzen, additives gaußverteiltes Rauschen zu unterdrücken [71], [21].

Kumulanten von multivariaten Zufallsvariablen werden auch als Kreuzkumulanten bezeichnet [47]. Sie müssen bei mittelwertfreien, statistisch unabhängigen Signalen für jede Ordnung verschwinden. Dabei reicht die alleinige Betrachtung der zweiten Ordnung allerdings nicht aus, da so lediglich die Signaldekorrelation, nicht aber statistische Unabhängigkeit gemessen werden kann. Aus diesem Grund muß auf Statistik höherer Ordnung zurückgegriffen werden.

Die Kumulante vierter Ordnung ist entsprechend Gleichung (4.79) von Seite 59 für mittelwertfreie Signale gegeben als

$$\begin{aligned} \text{cum}(x_1, x_2, x_3, x_4) &= E[x_1 x_2 x_3 x_4] - E[x_1 x_2] \cdot E[x_3 x_4] \\ &\quad - E[x_1 x_3] \cdot E[x_2 x_4] - E[x_1 x_4] \cdot E[x_2 x_3]. \end{aligned}$$

Für zwei komplexwertige Signale x_1 und x_2 kann die Berechnung der Kreuzkumulanten vierter Ordnung auf verschiedene Arten erfolgen. Aufgrund vielfältiger Symmetrien, wie

$$\begin{aligned} \text{cum}(x_1, x_2^*, x_3, x_4^*) &= \text{cum}(x_1, x_4^*, x_3, x_2^*) = \text{cum}(x_3, x_2^*, x_1, x_4^*) \\ &= \text{cum}(x_3, x_4^*, x_1, x_2^*) = \text{cum}^*(x_1, x_2^*, x_3, x_4^*) = \text{cum}^*(x_4, x_1^*, x_2, x_3^*) \\ &= \text{cum}^*(x_2, x_3^*, x_4, x_1^*) = \text{cum}^*(x_4, x_3^*, x_2, x_1^*), \end{aligned}$$

existieren jedoch nur drei betragsmäßig verschiedene Kreuzkumulanten vierter Ordnung für zwei komplexwertige Signale, die wie folgt definiert sind:

$$\text{cum}_{13}(x_1, x_2) = \text{cum}(x_1, x_2^*, x_2, x_2^*) \quad (5.15)$$

$$\text{cum}_{31}(x_1, x_2) = \text{cum}(x_1, x_1^*, x_1, x_2^*) \quad (5.16)$$

$$\text{cum}_{22}(x_1, x_2) = \text{cum}(x_1, x_1^*, x_2, x_2^*). \quad (5.17)$$

Es läßt sich zeigen [65], daß die Verwendung der in Gleichung (5.15) bis (5.17) definierten Kreuzkumulanten als ICA-Kostenfunktion bei unterschiedlichen Vorzeichen der Kurtosis der Quellsignale $\kappa(s_i)$ zu einer falschen Lösungen führen kann. Für gleiche Vorzeichen liefert die reellwertige, in Gleichung (5.17) definierte, Kreuzkumulante jedoch eindeutige Lösungen, wie aus der nachfolgenden Betrachtung ersichtlich wird.

Setzt man ein normiertes Mischungsmodell voraus, berechnet sich $\text{cum}_{22}(x_1, x_2)$, die Kreuzkumulante der Sensorsignale, zu

$$\begin{aligned} \text{cum}_{22}(x_1, x_2) &= \text{cum}(x_1, x_1^*, x_2, x_2^*) \\ &= \text{cum}\left((s_1 + a_{12}s_2), (s_1 + a_{12}s_2)^*, (a_{21}s_1 + s_2), (a_{21}s_1 + s_2)^*\right) \\ &= \text{cum}\left((s_1), (s_1)^*, (a_{21}s_1), (a_{21}s_1)^*\right) + \text{cum}\left((a_{12}s_2), (a_{12}s_2)^*, (s_2), (s_2)^*,\right) \\ &= a_{21}a_{21}^*\kappa(s_1) + a_{12}a_{12}^*\kappa(s_2) \\ &= |a_{21}|^2\kappa(s_1) + |a_{12}|^2\kappa(s_2). \end{aligned} \quad (5.18)$$

Definiert man das Produkt aus Mischungs- und Entmischungsmatrix als

$$\mathbf{M} = \mathbf{W} \cdot \mathbf{A}, \quad (5.19)$$

und betrachtet man die Abhängigkeit der Ausgangs- von den Quellsignalen als

$$\mathbf{y} = \mathbf{M} \cdot \mathbf{s}, \quad (5.20)$$

kann in äquivalenter Art und Weise zu Gleichung (5.18) die Abhängigkeit der Kreuzkumulante der Ausgangssignale von der Kurtosis der Quellsignale und der Matrix \mathbf{M}

beschrieben werden

$$\begin{aligned}
 \text{cum}_{22}(y_1, y_2) &= \text{cum}(y_1, y_1^*, y_2, y_2^*) \\
 &= \text{cum}\left((m_{11}s_1 + m_{12}s_2), (m_{11}s_1 + m_{12}s_2)^*, (m_{21}s_1 + m_{22}s_2), (m_{21}s_1 + m_{22}s_2)^*\right) \\
 &= \text{cum}\left((m_{11}s_1), (m_{11}s_1)^*, (m_{21}s_1), (m_{21}s_1)^*\right) \\
 &\quad + \text{cum}\left((m_{12}s_2), (m_{12}s_2)^*, (m_{22}s_2), (m_{22}s_2)^*\right) \\
 &= |m_{11}|^2|m_{21}|^2\kappa(s_1) + |m_{12}|^2|m_{22}|^2\kappa(s_2).
 \end{aligned} \tag{5.21}$$

Das Nullsetzen von $\text{cum}_{22}(y_1, y_2)$ liefert bei gleichem Vorzeichen von $\kappa(s_1)$ und $\kappa(s_2)$ die vier Lösungen

$$m_{11} = m_{12} = 0 \tag{5.22}$$

$$m_{11} = m_{22} = 0 \tag{5.23}$$

$$m_{21} = m_{12} = 0 \tag{5.24}$$

$$m_{21} = m_{22} = 0. \tag{5.25}$$

Dabei sind Gleichung (5.22) und (5.25) triviale Lösungen, da in diesem Fall jeweils eine gesamte Zeile der Matrix \mathbf{M} gleich Null ist und somit das entsprechende Ausgangssignal verschwindet. Gleichung (5.24) stellt die Lösung des ICA-Problems dar, da die Nebendiagonalelemente der Matrix \mathbf{M} verschwinden, das Produkt aus Mischungs- und Entmischungsmatrix also bis auf eine mögliche Skalierung einer Einheitsmatrix entspricht. Auch Gleichung (5.23) stellt eine Lösung des ICA-Problems mit möglicherweise skalierten Quellsignalen dar. Diese sind jedoch vertauscht, da die Hauptdiagonalelemente der Matrix \mathbf{M} in diesem Fall gleich Null sind.

Damit ist gezeigt, daß die in Gleichung (5.17) definierte Kreuzkumulante zweier komplexer Signale für Quellen mit gleichem Vorzeichen der Kurtosis als ICA-Kostenfunktion geeignet ist. Weist die Kurtosis der Quellsignale unterschiedliche Vorzeichen auf, kann durch Prewhitening der Sensorsignale sichergestellt werden, daß keine falsche Lösung gefunden wird [65]. Gleichzeitig muß man dabei aber neben einem leicht erhöhten Rechenaufwand in Kauf nehmen, daß die Interpretierbarkeit der Elemente der Entmischungsmatrix als blickrichtungsabhängige Phasenfaktoren im allgemeinen verloren geht. Hauptsächlich aus letzterem Grund und weil beim Anwendungsfall Sprachsignalverarbeitung die Quellsignale ausschließlich eine positive Kurtosis aufweisen, wird nachfolgend auf ein Prewhitening verzichtet.

Eine Vereinfachung bei der Berechnung der Kreuzkumulante vierter Ordnung ergibt sich durch Normierung der nach Gleichung (5.7) berechneten Ausgangssignale auf Einheitsvarianz

$$\tilde{Y}_1(\omega) = \frac{Y_1(\omega)}{\sigma_{Y_1}} = \frac{Y_1(\omega)}{\sqrt{\sigma_{Y_1}^2}} = \frac{Y_1(\omega)}{\sqrt{E[Y_1(\omega)Y_1^*(\omega)]}} \quad (5.26)$$

$$\tilde{Y}_2(\omega) = \frac{Y_2(\omega)}{\sigma_{Y_2}} = \frac{Y_2(\omega)}{\sqrt{\sigma_{Y_2}^2}} = \frac{Y_2(\omega)}{\sqrt{E[Y_2(\omega)Y_2^*(\omega)]}}. \quad (5.27)$$

In diesem Fall bleibt das Produkt $E[\tilde{Y}_1\tilde{Y}_1^*] \cdot E[\tilde{Y}_2\tilde{Y}_2^*]$ konstant gleich eins und braucht folglich bei der Optimierung der Kostenfunktion nicht weiter betrachtet zu werden. Damit ergibt sich die modifizierte Kreuzkumulante, die als Maß für die statistische Unabhängigkeit der Ausgangssignale dient, wie folgt:

$$\widehat{\text{cum}}_{22}(\tilde{Y}_1, \tilde{Y}_2) = E[\tilde{Y}_1 \cdot \tilde{Y}_1^* \cdot \tilde{Y}_2 \cdot \tilde{Y}_2^*] - E[\tilde{Y}_1\tilde{Y}_2^*] \cdot E[\tilde{Y}_2\tilde{Y}_1^*] - E[\tilde{Y}_1\tilde{Y}_2] \cdot E[\tilde{Y}_1^*\tilde{Y}_2^*]. \quad (5.28)$$

Für den Anwendungsfall Sprachsignalverarbeitung, bei dem beide Quellen eine positive Kurtosis aufweisen, würde schon die Minimierung von $\widehat{\text{cum}}_{22}$ zur richtigen Lösung führen. Für den allgemeineren Fall einer unbekanntem aber gleichen Kurtosis der Quellsignale muß die Kostenfunktion jedoch als Betrag der Kreuzkumulante vierter Ordnung definiert werden

$$\mathcal{K} = |\widehat{\text{cum}}_{22}(\tilde{Y}_1, \tilde{Y}_2)|. \quad (5.29)$$

5.2.4 Optimierung

Ziel der Optimierung ist es, die in Gleichung (5.29) definierte Kostenfunktion zu minimieren. Die dafür notwendige Berechnung der Ausgangssignale Y_1 und Y_2 kann in vereinfachter Form durch die Anwendung einer modifizierten Entmischungsmatrix $\tilde{\mathbf{W}}$ erfolgen, die sich lediglich in einem Skalierungsfaktor von der in Gleichung (5.6) definierten Inversen der Mischungsmatrix unterscheidet. Ohne Berücksichtigung von $\frac{1}{\det(\mathbf{A})}$ ergibt sie sich zu

$$\tilde{\mathbf{W}} = \begin{bmatrix} -e_2 & 1 \\ e_1 & -1 \end{bmatrix}. \quad (5.30)$$

Damit folgt für die beiden Ausgangssignale in jedem Frequenzband:

$$Y_1 = -e_2 X_1 + X_2 \quad (5.31)$$

$$Y_2 = e_1 X_1 - X_2. \quad (5.32)$$

Mit

$$\mathbf{Y} = \tilde{\mathbf{W}} \cdot \mathbf{X} \quad (5.33)$$

$$= \tilde{\mathbf{W}} \cdot \mathbf{A} \cdot \mathbf{S} \quad (5.34)$$

$$= \begin{bmatrix} -e_2 & 1 \\ e_1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ e_1 & e_2 \end{bmatrix} \cdot \mathbf{S} \quad (5.35)$$

$$= \begin{bmatrix} e_1 - e_2 & 0 \\ 0 & e_1 - e_2 \end{bmatrix} \cdot \mathbf{S} \quad (5.36)$$

ist die Unabhängigkeit der Ausgangssignale gewährleistet, wenn nach Voraussetzung die Quellsignale statistisch unabhängig sind. Zusätzlich sind die entmischten Signale im Vergleich zu den Quellsignalen mit dem Faktor $(e_1 - e_2)$ skaliert.

Für die Minimierung der Kostenfunktion mittels Gradientenabstiegsverfahren müssen zunächst die partiellen Ableitungen nach den konjugiert komplexen Parametern e_1^* und e_2^* gebildet werden [18]. Die Berücksichtigung der betragsmäßigen Minimierung der Kostenfunktion erfolgt in einfacher Weise durch Integration des Vorzeichens der Kostenfunktion in die partiellen Ableitungen. Aus Gründen der Übersichtlichkeit ist die Gradientenberechnung in Anhang A ab Seite 113 ausgeführt. Die beiden partiellen Ableitungen der Kostenfunktion ergeben sich zu

$$\begin{aligned} \frac{\partial \mathcal{K}}{\partial e_1^*} = \frac{\text{sign}(\mathcal{K})}{\sigma_{Y_1}^2 \sigma_{Y_2}^4} \cdot & \left(E[Y_1 Y_1^* Y_2 X_1^*] \sigma_{Y_2}^2 - E[Y_1 Y_1^* Y_2 Y_2^* E[Y_2 X_1^*]] - E[Y_1 X_1^*] E[Y_1^* Y_2] \sigma_{Y_2}^2 \right. \\ & \left. + \left| E[Y_1 Y_2^*] \right|^2 E[Y_2 X_1^*] - E[Y_1 Y_2] E[Y_1^* X_1^*] \sigma_{Y_2}^2 + \left| E[Y_1 Y_2] \right|^2 E[Y_2 X_1^*] \right) \end{aligned} \quad (5.37)$$

und

$$\begin{aligned} \frac{\partial \mathcal{K}}{\partial e_2^*} = \frac{\text{sign}(\mathcal{K})}{\sigma_{Y_1}^4 \sigma_{Y_2}^2} \cdot & \left(E[Y_1 Y_1^* Y_2 Y_2^* E[Y_1 X_1^*]] - E[Y_1 Y_2 Y_2^* X_1^*] \sigma_{Y_1}^2 - \left| E[Y_1 Y_2^*] \right|^2 E[Y_1 X_1^*] \right. \\ & \left. + E[Y_1 Y_2^*] E[X_1^* Y_2] \sigma_{Y_1}^2 - \left| E[Y_1 Y_2] \right|^2 E[Y_1 X_1^*] + E[Y_1 Y_2] E[Y_2^* X_1^*] \sigma_{Y_1}^2 \right). \end{aligned} \quad (5.38)$$

Da man an den optimalen Werten für φ_1 und φ_2 interessiert ist, ist der Gradient nach e_1 bzw. e_2 in dieser Form noch nicht geeignet für die Bestimmung des Minimums. Aufgrund der durch die Parametrisierung nach Gleichung (5.1) auf Seite 73 vorgegeben Beschränkung der Elemente der Entmischungsmatrix auf den Einheitskreis, liefert die Projektion

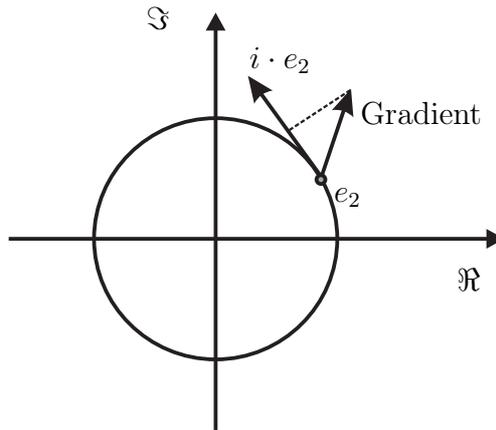


Abbildung 5.2: Projektion des Gradientenvektors auf den Einheitskreis

des Gradientenvektors auf die Tangente an den Einheitskreis den richtigen Wert für φ , wie anhand von Abbildung 5.2 verdeutlicht werden soll.

Die Projektion kann als Skalarprodukt zwischen Gradientenvektor und Tangentenvektor mit positiver φ -Richtung im jeweiligen Punkt e_1 bzw. e_2 berechnet werden

$$\frac{\partial \mathcal{K}}{\partial \varphi_1} = \left[\Re(\nabla_{e_1^*}) \quad \Im(\nabla_{e_1^*}) \right] \cdot \left[\Re(-ie_1) \quad \Im(-ie_1) \right]^T \quad (5.39)$$

$$\frac{\partial \mathcal{K}}{\partial \varphi_2} = \left[\Re(\nabla_{e_2^*}) \quad \Im(\nabla_{e_2^*}) \right] \cdot \left[\Re(ie_2) \quad \Im(ie_2) \right]^T. \quad (5.40)$$

Um die Optimierung weiter zu vereinfachen und eine realzeitfähige Implementierung zu ermöglichen (siehe Abschnitt 5.2.5.1), ist es weiterhin notwendig, den Algorithmus in aufwandsreduzierter Form zu formulieren, so daß die Abhängigkeit von den Ausgangssignalen in eine Abhängigkeit von den Sensorsignalen umgewandelt wird. Das heißt, die Auswertung der Kostenfunktion und des Gradienten soll nicht anhand der Ausgangssignale $Y_i(\omega)$, sondern durch Anwendung der in jedem Zwischenschritt neu geschätzten Parameter e_1 und e_2 auf die Meßsignalstatistik erfolgen.

Zu diesem Zweck ist es nötig, die in Gleichung (5.31) und (5.32) aufgestellten Vorschriften zur Bildung der Ausgangssignale in die Berechnung der Kostenfunktion und der Gradienten einzusetzen. Diese Berechnungen sind der Übersichtlichkeit halber in Anhang B ab Seite 117 für die Berechnung der aufwandsreduzierten Kostenfunktion bzw. ab Seite 120 für die Berechnung des aufwandsreduzierten Gradienten ausgeführt. Es ergeben sich insgesamt 12 verschiedene Erwartungswerte ν_1 bis ν_{12} , mit deren Hilfe sich die nötige

Messsignalstatistik beschreiben läßt

$$\begin{aligned}
 \nu_1 &= E[X_1^2] & \nu_2 &= E[X_2^2] \\
 \nu_3 &= E[|X_1|^2] & \nu_4 &= E[|X_2|^2] \\
 \nu_5 &= E[X_1 X_2] & \nu_6 &= E[X_1 X_2^*] \\
 \nu_7 &= E[X_1^2 X_2^{*2}] & \nu_8 &= E[|X_1|^2 |X_2|^2] \\
 \nu_9 &= E[X_1 X_2^* |X_1|^2] & \nu_{10} &= E[X_1 X_2^* |X_2|^2] \\
 \nu_{11} &= E[|X_1|^4] & \nu_{12} &= E[|X_2|^4].
 \end{aligned}$$

Damit nimmt die Kostenfunktion $\mathcal{K} = \mathcal{K}(e_1, e_2, X_1, X_2)$ die Form

$$\begin{aligned}
 \mathcal{K} &= \left| \frac{2\Re(e_1 e_2 \nu_7) + 2\Re(e_1 e_2^*) \nu_8 - 2\Re(e_1 \nu_9) - 2\Re(e_2 \nu_9) - 2\Re(e_2 \nu_{10}) - 2\Re(e_1 \nu_{10})}{(\nu_3 - 2\Re(e_2 \nu_6) + \nu_4) \cdot (\nu_3 - 2\Re(e_1 \nu_6) + \nu_4)} \right. \\
 &\quad \left. + \frac{2\nu_8 + \nu_{11} + \nu_{12} - \left| -e_1^* e_2 \nu_3 + e_2 \nu_6 + e_1^* \nu_6^* - \nu_4 \right|^2 - \left| -e_1 e_2 \nu_1 + e_2 \nu_5 + e_1 \nu_5 - \nu_2 \right|^2}{(\nu_3 - 2\Re(e_2 \nu_6) + \nu_4) \cdot (\nu_3 - 2\Re(e_1 \nu_6) + \nu_4)} \right| \quad (5.41)
 \end{aligned}$$

an. Für die Vorschrift des Gradienten sei auf Anhang B.2 verwiesen.

Die Minimierung der Kostenfunktion mittels einfachem Gradientenabstiegsverfahren bietet den Vorteil, eine effiziente Implementierung zu ermöglichen. Grundlage dafür ist die Ausnutzung der gegebenen Parametrisierung und die Normierung des Gradienten auf einen Betrag von Eins. Damit kann die bei Gradientenabstiegsverfahren kritische Wahl der Schrittweite anhand der natürlichen Vorgaben, d.h. anhand der tatsächlichen Gradzahlen erfolgen. Weitere Gründe, die für den Einsatz eines Gradientenabstiegsverfahrens sprechen sind, daß diese Methode durch die stete Einhaltung der Abstiegsbedingung sehr robust ist und pro Optimierungsschritt im Vergleich zu anderen Verfahren den wenigsten Rechenaufwand benötigt [72].

Das Schema aus Abbildung 5.3 faßt den Optimierungsalgorithmus zusammen, bei dem die im ersten Durchlauf maximale Schrittweite stufenweise verkleinert wird, bis man sich der Lösung auf eine Mindestgradzahl genähert hat. Diese Vorgehensweise führt zu einer durchschnittlichen Anzahl von 10 Optimierungsschritten pro Frequenzband bei einer Initialisierungsschrittweite von 20 Grad und sechsfacher stufenweiser Halbierung. D.h., das Optimum wird mit einer Genauigkeit von 0.3125 Grad gefunden.

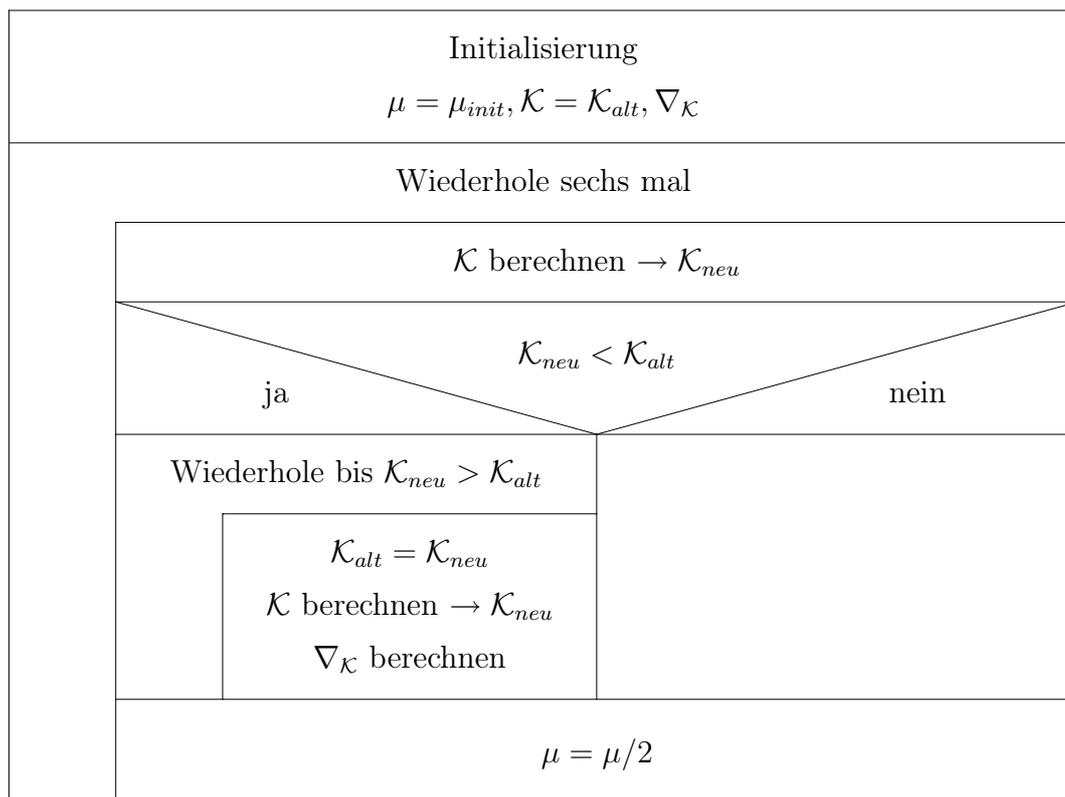


Abbildung 5.3: Schema des Optimierungsalgorithmus

5.2.5 Praktische Gesichtspunkte

Das vorgestellte Verfahren des frequenzvarianten Nullbeamforming findet die auszunul-lenden Richtungen mit Hilfe von Statistik höherer Ordnung, so daß die entstehenden Ausgangssignale statistisch unabhängig voneinander sind. Für die Realisierung dieses Verfahrens als praxistaugliches System sind einige zusätzliche Gesichtspunkte, wie z.B. Realzeitfähigkeit, Mikrofonabstand und Raumbedingungen zu berücksichtigen, die nach-folgend näher untersucht werden.

5.2.5.1 Realzeitfähigkeit

Bei der Implementierung des Verfahrens als realzeitfähiges System müssen entweder einzel-ne Datenblöcke benutzt oder eine rekursive Schätzung zur Bestimmung der statistischen Parameter durchgeführt werden. Da bei einer blockorientierten Arbeitsweise Unstetigkei-ten an den Blockgrenzen auftreten können, die in den Ausgangssignalen deutlich zu hören sind, bietet sich eine fortlaufenden Schätzung der Signalstatistik an. Ein dafür weit ver-breiteter Ansatz ist die rekursive Mittelung erster Ordnung, bei der ein zu schätzender

Parameter θ im aktuellen Durchgang k als

$$\theta(k) = \beta \cdot \theta(k-1) + (1 - \beta) \cdot \hat{\theta}(k) \quad (5.42)$$

bestimmt wird. Damit ist, in Abhängigkeit von der Lernrate β , eine kontinuierliche Adaption an sich verändernde Umgebungsbedingungen gegeben. Der entscheidende Vorteil im Vergleich zu blockorientierten Verfahren liegt jedoch im wesentlich geringeren Speicherbedarf, da keine langen Puffer vorgehalten werden müssen, sondern jeder zu schätzende Parameter nur einmal im Speicher abgelegt ist.

Um das frequenzvariante Nullbeamforming mit rekursiver Parameterschätzung realisieren zu können, müssen Kostenfunktion und Gradient, wie in Abschnitt 5.2.4 beschrieben, in Abhängigkeit von Eingangssignalen und Phasenfaktoren formuliert sein. Dies ist notwendig, da andernfalls in einem Optimierungsschritt zur Schätzung der Kostenfunktion ein Zugriff auf das gesamte Ausgangssignal ermöglicht werden müßte. In dieser Form ließe sich damit nur eine blockorientierte Vorgehensweise umsetzen.

Ein weiterer Vorteil des Verfahrens mit rekursiv geschätzter Signalstatistik ist, daß für die Berechnung von Kostenfunktion und Gradient wesentlich weniger Rechenaufwand nötig ist. Besonders deutlich wird der Unterschied im Vergleich zum Batch-Verfahren, bei dem das komplette Meßsignal in einem Durchgang betrachtet wird. Nimmt man z.B. eine gesamte Anzahl von 1000 Frames an¹, müssen allein zur Bildung des bei der Normierung notwendigen Erwartungswertes $E[Y_i Y_i^*]$ in jedem Frequenzband sowohl 1000 komplexwertige Multiplikationen als auch 999 Additionen gerechnet werden. Im Vergleich dazu kommt man bei rekursiver Schätzung pro Optimierungsschritt und Frequenzband mit 16 komplexwertigen Multiplikationen und fünf Additionen aus, wie sich am Beispiel von $E[Y_1 Y_1^*]$ nachvollziehen läßt

$$\begin{aligned} E[Y_1 Y_1^*] &= E[(-e_2 X_1 + X_2)(-e_2^* X_1^* + X_2^*)] \\ &= |e_2|^2 E[|X_1|^2] - 2\Re(e_2 E[X_1 X_2^*]) + E[|X_2|^2]. \end{aligned}$$

Unter der Annahme von sich langsam bewegenden Schallquellen dürfen φ_1 und φ_2 in aufeinanderfolgenden Schritten nur wenig geändert werden. Deshalb werden diese Größen ebenfalls rekursiv bestimmt. D.h., ein neu berechneter Winkel geht nur zu einem bestimmten Teil in den aktuell gültigen Wert für das jeweilige φ ein. Diese Vorgehensweise erhöht die Robustheit des Algorithmus gegenüber statistischen Ausreißern und möglichen Permutationsproblemen [39].

¹ ≈ 11.7 s bei $f_s = 11$ kHz, FFT-Länge $L = 512$, Versatz $V = 128$

5.2.5.2 Raumbedingungen

Die akustischen Eigenschaften eines Raumes werden im Wesentlichen durch die Impulsantwort charakterisiert, die die Filterung eines Signals innerhalb eines Raumes beschreibt. Eine Möglichkeit, akustische Bedingungen miteinander zu vergleichen, ist durch die Bewertung der Länge der Impulsantwort in Form der Nachhallzeit gegeben. Das ist diejenige Zeit, die vergeht, bis die Energie der Impulsantwort auf ein Millionstel ihres Anfangswertes, also um 60 dB, abgeklungen ist [57]. Aus diesem Grund wird die Nachhallzeit oft mit τ_{60} bezeichnet. Je länger sie ist, desto schwieriger sind die akustischen Bedingungen im jeweiligen Raum und desto schlechter funktionieren akustische Beamformer und Verfahren der konvolutiven Quellentrennung.

Um den Einfluß der akustischen Eigenschaften eines Raumes zu verdeutlichen, soll die tatsächliche Einfallsrichtung eines Schallsignals bei verschiedenen Konfigurationen untersucht werden. Grundlage dafür ist das Simulationsprogramm *room* [103], mit dessen Hilfe sich Raumimpulsantworten in Abhängigkeit von den Reflexionskoeffizienten der Wände sowie von Schallquellen- und Sensorposition ermitteln lassen. Abbildung 5.4 zeigt den für die nachfolgenden Simulationen zugrunde gelegten 50 m³ großen Raum. Die Schallquelle befindet sich in einem Abstand von 1.5 m zur linken und 1 m zur vorderen Wand, und die Mikrofone sind in gleicher Höhe im Abstand d mit einer Entfernung von 1.4 m zur Schallquelle platziert.

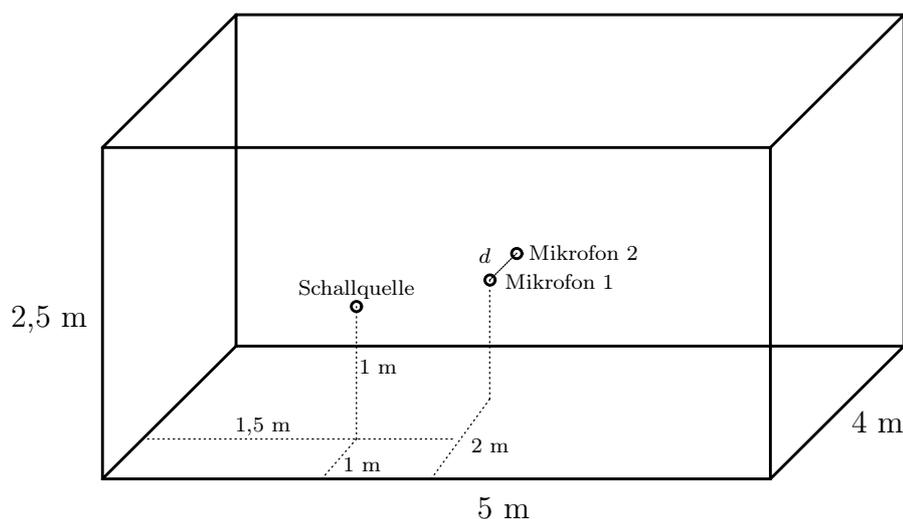


Abbildung 5.4: Simulierter Raum

Um eine Vorstellung von den Auswirkungen verschiedener Raumbedingungen zu bekom-

men, können bei bekannter Übertragungsfunktion die tatsächlichen, frequenzabhängigen Schalleinfallrichtungen bestimmt werden. D.h., die Phasenanteile der Übertragungsfunktion zwischen Quelle und Mikrofonen werden mit Hilfe der Array-Antwort des Fernfeld-Beamformingmodells aus Abschnitt 3.1.3 ausgewertet. Wählt man ein Nullbeamformingmodell, läßt sich aus dem Beampattern die Lage der in jedem Frequenzband ausgebildeten Null und damit die tatsächliche Schalleinfallrichtung sehr gut bestimmen.

Für das nachfolgende Beispiel wurden die Reflexionskoeffizienten aller Wände auf den gleichen Wert gesetzt und dieser stufenweise in den Schritten $\rho = 0.2$, $\rho = 0.5$ und $\rho = 0.8$ variiert. Im simulierten Raum ergeben sich dadurch Nachhallzeiten von ca. 112 ms, 192 ms und 480 ms, da aufgrund der höheren Anzahl an Reflexionen auch die Länge der Impulsantworten anwächst. Abbildung 5.5 zeigt die durch die jeweilige Übertragungsfunktion verursachte Streuung der Einfallrichtung im Bereich bis 1700 Hz.

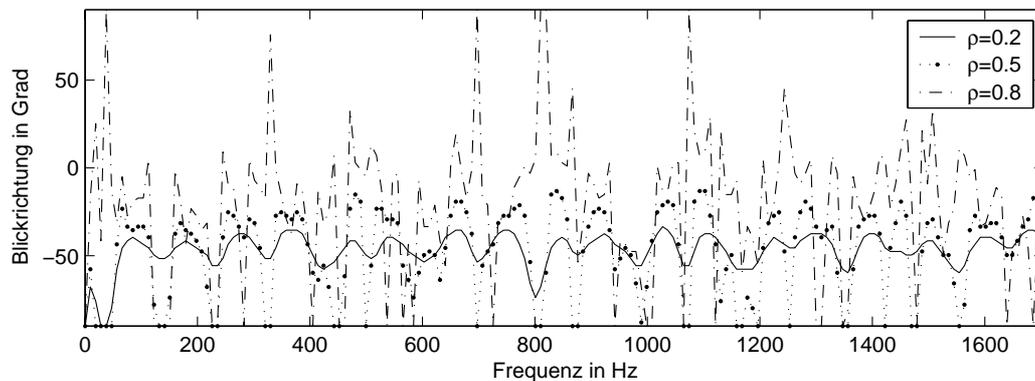


Abbildung 5.5: Streuung der Einfallswinkel über der Frequenz bei verschiedenen Raumkonfigurationen

Offensichtlich, und wie nicht anders zu erwarten, geht mit zunehmender Länge der Raumimpulsantwort bzw. der Nachhallzeit die Eindeutigkeit der Einfallrichtung zunehmend verloren. Bei kurzer Nachhallzeit liegen die Schwankungen nah bei der durch die Anordnung von Schallquelle und Mikrofonarray vorgegebenen geometrischen Einfallrichtung von -45° . Nimmt die Nachhallzeit zu, weicht die tatsächliche Einfallrichtung immer weiter von der geometrischen ab, so daß Streuungen über den gesamten Blickrichtungsbereich möglich sind.

An diesem Beispiel wird deutlich, daß die erreichbare Störsignalunterdrückung stark von der akustischen Umgebung abhängig ist. Während im Freifeld bzw. bei einem Reflexionskoeffizienten gleich Null eine über alle Frequenzbänder konstante Ausnullungsrichtung für die Ausblendung eines Störers noch ausreichend ist, muß in realen akustischen Umgebungen die Lage der Null in jedem Frequenzband adaptiert werden.

Ein weiterer, die ideale geometrische Einfallsrichtung verfälschender Faktor, ist durch die Abweichung vom Modell der planaren Schallwellen des Fernfeldes bei kürzeren Abständen zwischen Quelle und Mikrofon gegeben. Insbesondere für niedrige Frequenzen müßte der korrekte Zeitversatz mit einem Nahfeldmodell bestimmt werden. Da die verursachten Abweichungen im Vergleich zum Phasenversatz durch die Übertragungsfunktion vernachlässigbar sind, kann auf die Anwendung des Nahfeld-Modells jedoch verzichtet werden.

5.2.5.3 Mikrofonabstand

Die Wahl des Mikrofonabstandes ist von entscheidender Bedeutung für eine robuste Funktionsweise des vorgestellten Verfahrens. So ist die eindeutige Zuordnung der Nullrichtungen nur möglich, wenn der maximale Mikrofonabstand unterhalb der Grenze für räumliches Aliasing liegt (vgl. Abschnitt 3.1.3). Oberhalb treten die in jedem Frequenzband bestimmten Blickrichtungen mehrfach auf und werden in die Bereiche des jeweils anderen Signals hineingespiegelt.

Ein weiterer Grund für einen möglichst kleinen Mikrofonabstand liegt in der verbesserten Glattheit der Kreuzfilter des normierten konvolutiven Mischungsmodelles. Diese Filter beschreiben die Ähnlichkeit der Signale an den Mikrofonen und sind somit im Zeitbereich um so kürzer, je näher beide Mikrofone beieinander platziert sind. Damit ist im Frequenzbereich die Streuung der Nullen um die Einfallsrichtung ebenfalls verringert. Mit der bereits im vorigen Abschnitt eingeführten Interpretation der Impulsantwort als Einfallsrichtung soll diese Tatsache verdeutlicht werden. Abbildung 5.6 zeigt die Beträge der Beampattern der Impulsantworten für einen Mikrofonabstand von einem bzw. 10 Zentimeter.

Der Wahl eines möglichst kleinen Mikrofonabstandes sind in der Realität jedoch Grenzen gesetzt. Zum einen läßt sich aufgrund der Größe der Mikrofone selbst nur ein bestimmter Mindestabstand erreichen. Zum anderen ist ein sinnvoller Mindestabstand durch die notwendige Begrenzung der Verstärkung unkorrelierten Rauschens gegeben.

Während das Dimensionsproblem durch den Einsatz von Siliziummikrofonen (siehe z.B. [79]) umgangen werden könnte, ist das starke Ansteigen der Rauschverstärkung strukturell bedingt und kann nicht verhindert werden. Die komplexen, auf den Einheitskreis beschränkten Phasenfaktoren unterscheiden sich bei niedrigen Frequenzen kaum für verschiedene Blickrichtungen. Daraus folgt, daß das Platzieren einer Null für das Auslöschen eines Signals aus einer bestimmten Richtung sich auf den gesamten Bereich möglicher Einfallsrichtungen auswirkt. Damit werden alle Signale, unabhängig von ihrer speziellen

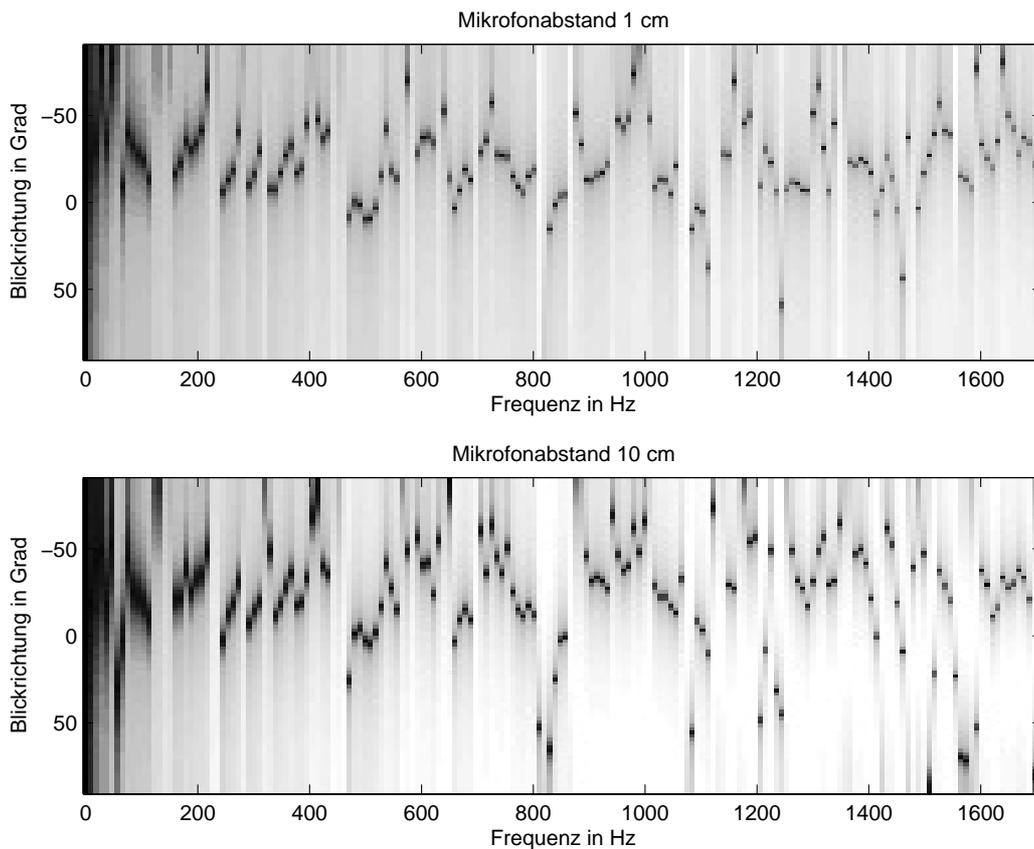


Abbildung 5.6: Streuung der Einfallrichtung in Abhängigkeit vom Mikrofonabstand

Einfallrichtung, sehr stark gedämpft. Diese Dämpfung muß dann betragsmäßig durch die Filterkoeffizienten wieder kompensiert werden.

Zur Verdeutlichung der Ähnlichkeit der Phasenfaktoren soll ihre Lage in der komplexen Ebene betrachtet werden. Dazu sei ein Phasenfaktor e_i dargestellt als

$$e_i = |e_i| \cdot \exp(i\angle e_i). \quad (5.43)$$

Da die Phasenfaktoren per Definition einen Betrag von Eins aufweisen, beschränkt sich die Auswertung auf den Winkel $\angle e_i$, der parametrisiert ist durch

$$\angle e_i = -\omega \frac{d}{c} \sin(\phi). \quad (5.44)$$

In Abbildung 5.7 auf der nächsten Seite ist dieser Winkel über dem möglichen Blickrichtungsbereich für verschiedene Frequenzen dargestellt. Dabei wird deutlich, daß er für kleine Frequenzen, wie z.B. 100 Hz, für alle Blickrichtungen weitgehend konstant bleibt. Für höhere Frequenzen ist die Unterscheidbarkeit deutlich besser, aber nicht über den gesamten Bereich gleich.

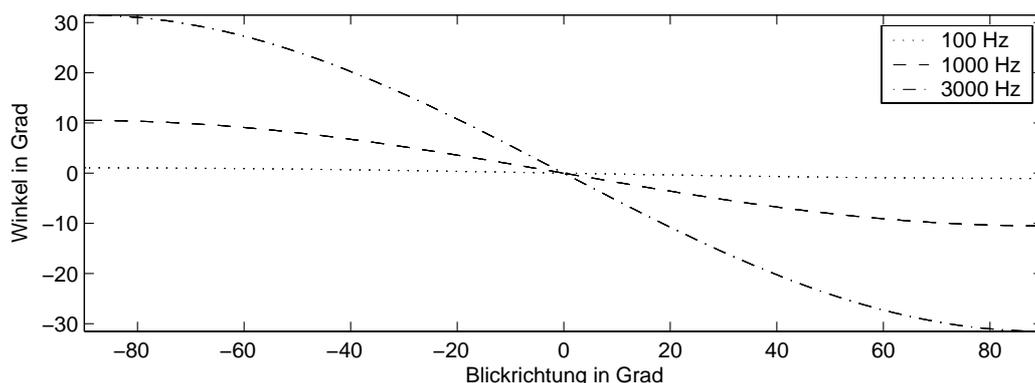


Abbildung 5.7: Winkel der Phasenfaktoren bei Mikrofonabstand $d=1\text{cm}$ und verschiedenen Frequenzen

Um eine Aussage über die tatsächlich zu erwartende Rauschverstärkung treffen zu können, soll zunächst das Verfahren mit vollständiger Entmischungsmatrix, d.h. unter Berücksichtigung des Vorfaktors $\frac{1}{\det(\mathbf{A})}$, analysiert werden. Der aus der ersten Zeile der Entmischungsmatrix gebildete Nullbeamformer, mit dem Koeffizientenvektor \mathbf{w}_1 , weist die Einträge

$$\mathbf{w}_1 = \begin{bmatrix} \frac{-e_2}{e_1 - e_2} & \frac{1}{e_1 - e_2} \end{bmatrix} \quad (5.45)$$

auf. Die Rauschverstärkung des Beamformers berechnet sich als Betragsquadrat des Koeffizientenvektors (siehe Abschnitt 3.2.2.3 auf Seite 24)

$$\mathbf{w}_1 \mathbf{w}_1^H = \begin{bmatrix} \frac{-e_2}{e_1 - e_2} & \frac{1}{e_1 - e_2} \end{bmatrix} \cdot \begin{bmatrix} \frac{-e_2}{e_1 - e_2} & \frac{1}{e_1 - e_2} \end{bmatrix}^H \quad (5.46)$$

$$= \frac{|e_2|^2 + 1}{|e_1 - e_2|^2} \quad (5.47)$$

$$= \frac{2}{|e_1 - e_2|^2} \quad (5.48)$$

$$(5.49)$$

Für den aus der zweiten Zeile der Entmischungsmatrix gebildeten Nullbeamformer ergibt sich äquivalent

$$\mathbf{w}_2 \mathbf{w}_2^H = \frac{|e_2|^2 + 1}{|e_1 - e_2|^2} \quad (5.50)$$

$$= \frac{2}{|e_1 - e_2|^2} \quad (5.51)$$

Die für beide Nullbeamformer gleiche Rauschverstärkung ist damit in Abhängigkeit von der Frequenz und der Einfallsrichtung formuliert. Je ähnlicher sich die beiden Phasenfak-

toren sind, desto schlechter ist die Mischungsmatrix konditioniert und desto größer ist die daraus resultierende Rauschverstärkung. Die Ähnlichkeit der Phasenfaktoren nimmt mit abnehmender Frequenz (vgl. Abb. 5.7) und kleineren Unterschieden im Einfallswinkel zu.

Um die Größenverhältnisse zu verdeutlichen, ist in der nachfolgenden Abbildung die nach Gleichung (5.48) bzw. (5.51) berechnete frequenzabhängige Rauschverstärkung für einen Mikrofonabstand von einem Zentimeter und angenommenen Signaleinfallrichtungen von $\pm 45^\circ$ dargestellt. Es ist zu sehen, daß im unteren Nutzsinalbereich bis 500 Hz unkorreliertes Rauschen mit über 20 dB verstärkt wird. Bei enger zusammen liegenden Einfallrichtungen von z.B. $\pm 20^\circ$ werden Werte über 30 dB erreicht.

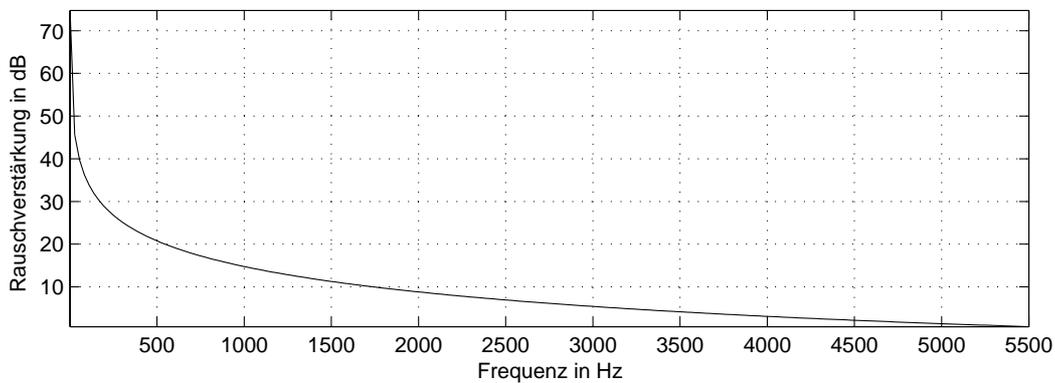


Abbildung 5.8: Rauschverstärkung des zweikanaligen Nullbeamformers bei Mikrofonabstand $d=1\text{cm}$

Da für praktische Anwendungen nur eine Rauschverstärkung von wenigen dB akzeptabel ist, kann das Verfahren in dieser Form nicht verwendet werden. Durch die Verwendung der modifizierten Entmischungsmatrix, wie in Abschnitt 5.2.4 vorgeschlagen, entfällt der für das starke Anwachsen der Rauschverstärkung verantwortliche Vorfaktor. Damit entfällt gleichzeitig die Skalierung der Sensorsignale, so daß die Rauschverstärkung für alle Frequenzen und Blickwinkel einen konstanten Wert von ca. 3 dB annimmt, wie am Beispiel von $\tilde{\mathbf{w}}_1$ deutlich wird

$$\tilde{\mathbf{w}}_1 \tilde{\mathbf{w}}_1^H = \begin{bmatrix} e_2 & -1 \end{bmatrix} \cdot \begin{bmatrix} e_2 & -1 \end{bmatrix}^H \quad (5.52)$$

$$= |e_2|^2 + 1 \quad (5.53)$$

$$= 2. \quad (5.54)$$

Mit Hilfe der modifizierten Entmischungsmatrix $\tilde{\mathbf{W}}$ läßt sich folglich das Problem der schlecht konditionierten Mischungsmatrix und der damit verbundenen übermäßigen Verstärkung unkorrelierten Rauschens umgehen. Wie in Abschnitt 5.2.4 erläutert, resultiert

diese Vorgehensweise in Ausgangssignalen, die sich durch den frequenz- und blickwinkelabhängigen Faktor $(e_1 - e_2)$ von den Quellsignalen unterscheiden. Dieser Faktor läßt sich als Hochpaßfilter mit Präemphase-Charakteristik interpretieren, dessen Betragsfrequenzgang, abgesehen von einem konstanten Offset, invers zur Rauschverstärkung aus Abbildung 5.8 verläuft. Die Ausgangssignale weisen dann wesentlich weniger Frequenzanteile unterhalb von 300 Hz auf, wodurch sich die Sprachverständlichkeit jedoch nicht verschlechtert.

5.2.5.4 Erweiterbarkeit auf mehrere Quellen

In der vorgestellten Form ist das Verfahren des frequenzvarianten Nullbeamforming für ein 2×2 -Szenario entwickelt, bei dem zwei Schallquellen von zwei Mikrofonen erfaßt werden. Die Erweiterung auf eine größere Anzahl an Quellen ist prinzipiell möglich, jedoch mit einem erheblichen Mehraufwand verbunden. Prinzipiell müssen bei adaptiven Beamforming-Verfahren für das Ausblenden von N Störsignalen $N+1$ Sensoren vorhanden sein. Ebenso ist bei den linearen Verfahren der blinden Quellentrennung mindestens die gleiche Anzahl an Sensoren nötig, wie Signalquellen vorhanden sind.

Folglich macht die Erweiterung auf mehr als zwei Quellen den Einsatz von genauso vielen Mikrofonen nötig, was bei vielen Anwendungen aus Kosten- und Platzgründen unerwünscht ist. Des weiteren steigt der Rechenaufwand sehr stark an, da für die Bestimmung der statistischen Unabhängigkeit der Ausgangssignale mehrere Kreuzkumulanten ausgewertet und im Verbund optimiert werden müssen.

Zu beachten ist weiterhin, daß die erreichbaren Trennungsergebnisse auch bei erfolgreicher Optimierung immer von den akustischen und geometrischen Gegebenheiten abhängen. Nur solange die einzelnen Quellen räumlich unterscheidbar sind, kann eine sinnvolle Zuordnung der Frequenzbänder zur entsprechenden Quelle erfolgen. Für Räume mit geringer Nachhallzeit und räumlich weit verteilten Quellen lassen sich drei bis vier Quellen sehr gut unterscheiden (siehe Abbildung 5.5). Bei noch mehr Quellen oder längerer Nachhallzeit reicht die alleinige räumliche Zuordnung der Frequenzbänder zu den Quellen nicht aus, und das damit entstehende Permutationsproblem muß unter Zuhilfenahme weiterer Kriterien gelöst werden.

5.2.5.5 Berücksichtigung des Betragsanteils

Bei der Herleitung des Algorithmus wurde das Fernfeld-Beamforming-Modell als Parametrisierung für die Elemente der Entmischungsmatrix zugrunde gelegt. Dies impliziert,

daß die Filterkoeffizienten e_1 und e_2 aus Gleichung (5.30) reine Phasenfaktoren sind, das Entmischungssystem also aus 2 Allpaßfiltern besteht.

Mit dieser Anordnung lassen sich sehr gute Trennungsergebnisse erzielen, da zum einen für das Auslöschung eines Signals die exakte Phaseninformation wichtiger als die exakte Amplitude ist. Zum anderen ist die Annahme einer ungefähr gleichen Amplitude zumindest bei kleinen Mikrofonabständen gerechtfertigt, wie durch Simulationen gezeigt werden kann. Abbildung 5.9 stellt die zu erwartenden Amplitudenunterschiede zwischen zwei Mikrofonen mit den Abständen 1 cm und 10 cm dar. Dabei ist der Simulation wieder der Raum und die Quellen- und Sensoranordnung aus Abbildung 5.4 zugrunde gelegt, wobei der Reflexionskoeffizient für alle Wände auf $\rho = 0.7$ gesetzt ist.

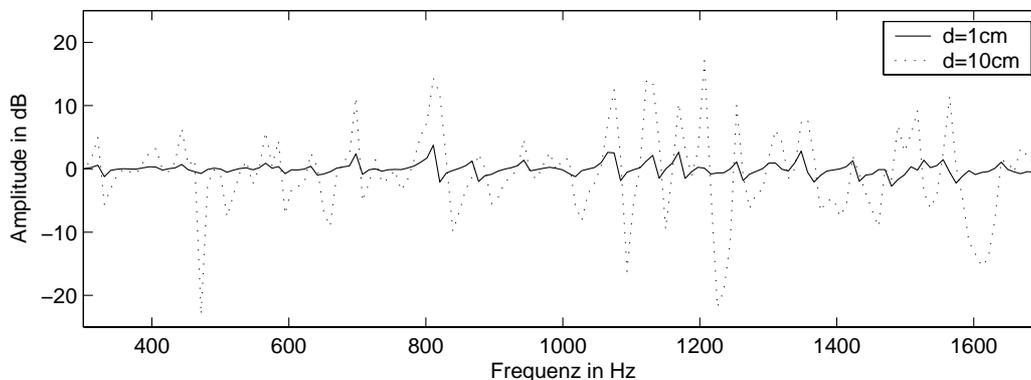


Abbildung 5.9: Amplitudenunterschiede in Abhängigkeit vom Mikrofonabstand

Ebenso wie der Phasenverlauf nimmt auch der Amplitudenverlauf einen umso glatteren Verlauf an, je kürzer die Nachhallzeit des Raumes ist. Für das freie Schallfeld bzw. bei einem allen Wänden gemeinsamen Reflexionskoeffizienten von $\rho = 0$ wäre die Amplitude über alle Frequenzen konstant gleich eins. Da diese idealen Verhältnisse in realen akustischen Umgebungen nicht anzutreffen sind, kann es vorteilhaft sein, für die Filterkoeffizienten $e_1(\omega)$ und $e_2(\omega)$ einen Betragsanteil zuzulassen. Damit läßt sich bei geeigneter Optimierung eine bessere Anpassung des Entmischungssystems an die Umgebungsbedingungen erreichen, woraus eine verbesserte Störsignaldämpfung und weniger Verzerrungen des Nutzsignals resultieren.

Für die zusätzliche Berücksichtigung der Beträge läßt sich die bereits definierte aufwandsreduzierte Kostenfunktion entsprechend Gleichung (B.12) und der aufwandsreduzierte Gradient aus (B.19) und (B.20) verwenden, da die Parameter e_1 und e_2 zunächst beliebige komplexe Größen darstellen. Lediglich die vorher zur Gradientenberechnung notwendige Projektion auf den Einheitskreis entfällt im Vergleich zum Verfahren mit reinen Phasenfaktoren.

Die größten Unterschiede zwischen beiden Verfahren entstehen bei der Optimierung, da Betrag und Phase nicht unabhängig voneinander sind und somit im Verbund optimiert werden müssen. Da sich für den Betragsanteil a priori kein begrenzter Suchraum angeben läßt, kann die effiziente Schrittweitensteuerung anhand der tatsächlichen Gradzahlen deshalb nicht verwendet werden. Aus diesem Grund erfolgt der Gradientenabstieg, ähnlich zum Verfahren nach Abbildung 5.3, mit einer willkürlichen Schrittweite, die stufenweise verkleinert wird, wodurch sich jedoch auch die Adaptionsgeschwindigkeit verkleinert.

Ein weiterer Nachteil dieses Verfahrens ist, daß während der Optimierung trotz geeigneter Initialisierung Vertauschungen zwischen den Frequenzbändern auftreten können und folglich eine zusätzliche Permutationskorrektur nötig ist. Diesen Nachteil weist das auf reinen Phasenfaktoren basierende Verfahren nicht auf, da bei räumlich getrennten Quellen und einer auf den Durchschnittswerten anderer Frequenzbänder basierenden Initialisierung Permutationen quasi ausgeschlossen sind.

Die Umstellung auf die unbeschränkte Optimierung bietet sich insbesondere dann an, wenn die in Abschnitt 5.2.5.3 beschriebene Hochpaßfilterung der Ausgangssignale nicht toleriert werden kann. Bei genügend großem Mikrofonabstand (> 2 cm) bewirken die unterschiedlichen Beträge der Koeffizienten eine verbesserte Konditionierung der Mischungsmatrix. Damit entschärft sich die Problematik der zu großen Rauschverstärkung und für die Entmischung kann in jedem Frequenzband die vollwertige Entmischungsmatrix \mathbf{W} verwendet werden. Für eine erhöhte Robustheit sollte der aus der Determinante der Mischungsmatrix gebildete Vorfaktor allerdings in jedem Frequenzband begrenzt werden, um einem möglicherweise zu großen Dynamikbereich der Entmischungsfiler vorzubeugen.

6 Ergebnisse und Diskussion

In diesem Kapitel soll die Leistungsfähigkeit des frequenzvarianten Nullbeamforming bei festen Quellsignalpositionen in Abhängigkeit von der jeweiligen akustischen Situation untersucht werden. Die in dieser Hinsicht relevanten Parameter sind der Mikrofonabstand, die Nachhallzeit und die räumliche Quellenverteilung. Im Hinblick auf eine realzeitfähige Implementierung ist zusätzlich die benötigte Rechenzeit dokumentiert.

6.1 Qualitätsmaße

Zur Beurteilung der Leistungsfähigkeit des Algorithmus ist zunächst ein entsprechendes Qualitätsmaß festzulegen. Für Verfahren der instantanen blinden Quellentrennung existieren verschiedene Bewertungskriterien, wie z.B. Signal-to-Noise-Ratio (SNR), Multichannel-Intersymbol-Interference (MISI) oder Mean-Squared-Error (MSE) [100]. Weiterhin ist es üblich, das Produkt aus Mischungs- und geschätzter Entmischungsmatrix im Hinblick auf die Abweichung von einer skalierten und permutierten Einheitsmatrix zu bewerten.

Für reale Raumaufnahmen konvolutiver Mischungen sind diese Verfahren jedoch nur bedingt geeignet. Die Bewertung der Abweichung von der Einheitsmatrix ist hier nicht möglich, da meist keine Informationen über die Mischung verfügbar sind. Lediglich für künstlich erzeugte Mischungen läßt sich das Verfahren benutzen, spiegelt jedoch nicht immer die tatsächliche Trennungsqualität wieder. So können Abweichungen von der Einheitsmatrix in Frequenzbändern mit niedriger Signalenergie zu einem verschlechterten Trennungsmaß führen, obwohl sie möglicherweise keine hörbaren Auswirkungen haben [82].

Aus diesem Grund sollten die Signalleistungen zur Beurteilung der erreichten Dämpfung berücksichtigt werden. Dazu ist es notwendig, die Quellen einzeln aufzuzeichnen, damit sich die berechneten Entmischungsfiler separat auf das jeweilige Signal anwenden lassen. Das für die Berechnung der Filter notwendige Summensignal kann durch Addition der Teilsignale erzeugt werden. Durch diese Vorgehensweise erreicht man eine maximal realistische Mischung, ohne daß Informationen, die zur Auswertung benötigt werden, verloren gehen [82].

6.1.1 Störsignaldämpfung

Mit Hilfe der einzeln aufgezeichneten Quellsignale läßt sich die Störsignaldämpfung als SIR (*signal to interference ratio*) angeben. Das SIR ist das Verhältnis der Leistungen von ungedämpftem Ausgangssignal zur Leistung der Summe der anderen Ausgangssignale und berechnet sich für Ausgang j als

$$SIR_j = 10 \log_{10} \left(\frac{E[y_{j,s_j}^2(t)]}{E[(\sum_{i \neq j} y_{j,s_i}(t))^2]} \right). \quad (6.1)$$

Dabei stellt $y_{j,s_i}(t)$ das Ausgangssignal j dar, wenn bei einer Einzelaufnahme nur Quelle i aktiv ist [81]. D.h., für $j = i$ sollte das Ausgangssignal keine Dämpfung im Vergleich zum Quellsignal aufweisen, und für $j \neq i$ sollte die Dämpfung maximal sein.

Für den im Rahmen dieser Arbeit betrachteten zweikanaligen Fall ergibt sich vereinfachend

$$SIR_j = 10 \log_{10} \left(\frac{E[y_{j,s_j}^2(t)]}{E[y_{j,s_i}^2(t)]} \right) \quad \text{mit } i \neq j. \quad (6.2)$$

Da man im allgemeinen nicht an den absoluten Werten, sondern an der Verbesserung des SIR interessiert ist, muß zusätzlich das Eingangs-SIR der Mikrofonssignale berücksichtigt werden [54]. Bei einzeln aufgezeichneten Mikrofonssignalen $x_{j,s_i}(t)$ ergibt sich die SIR-Verbesserung eines Kanals zu

$$SIR_j^+ = 10 \log_{10} \left(\frac{E[y_{j,s_j}^2(t)]}{E[y_{j,s_i}^2(t)]} \right) - 10 \log_{10} \left(\frac{E[x_{j,s_j}^2(t)]}{E[x_{j,s_i}^2(t)]} \right) \quad \text{mit } i \neq j. \quad (6.3)$$

Zur Bewertung der Ergebnisse im Hinblick auf die erreichte Störsignaldämpfung werden damit die beiden nachfolgend definierten SIR-Verbesserungen verwendet

$$SIR_1^+ = 10 \log_{10} \left(\frac{E[y_{1,s_1}^2]}{E[y_{1,s_2}^2]} \right) - 10 \log_{10} \left(\frac{E[x_{1,s_1}^2]}{E[x_{1,s_2}^2]} \right) \quad (6.4)$$

$$SIR_2^+ = 10 \log_{10} \left(\frac{E[y_{2,s_2}^2]}{E[y_{2,s_1}^2]} \right) - 10 \log_{10} \left(\frac{E[x_{2,s_2}^2]}{E[x_{2,s_1}^2]} \right). \quad (6.5)$$

Es ist jedoch zu beachten, daß diese SIR-Bewertung nur für rauschfreie Umgebungen genaue Ergebnisse liefert. Bei realen Raumaufnahmen ist das SNR (*signal to noise ratio*) stets begrenzt und damit die Genauigkeit der entsprechenden dB-Werte eingeschränkt. Die Bewertung stark verrauschter Signale, wie z.B. bei Fahrten im Automobil, würde zusätzliches Wissen über die einzelnen Rauschleistungen erfordern.

6.1.2 Nutzsinalstörung

Neben der Bewertung der Störsignalunterdrückung muß auch die durch den Algorithmus verursachte Nutzsinalverzerrung berücksichtigt werden. Bei der Arbeit mit realen akustischen Aufnahmen gestaltet sich die objektive Bewertung der Nutzsinalstörung jedoch als sehr kompliziert, da entsprechende Maßzahlen meist nicht in Relation zum subjektiven Höreindruck stehen. Insbesondere im Bereich der Sprachsignalverarbeitung stellt die Qualität des Signals eine mehrdimensionale Größe mit zahlreichen Aspekten dar [95]. Die Bewertung der Nutzsinalstörung kann deshalb nicht mit der Sprachqualität gleichgesetzt werden.

Eine Methode für die Bewertung der Nutzsinalstörung ist der direkte Vergleich von Ausgangs- und Originalsignal. Dieser Vergleich sollte, im Unterschied zur Bewertung der Trennungsqualität, nicht im Zeitbereich erfolgen, da aufgrund der eingesetzten Filter die Zeitsignale nicht direkt vergleichbar sind und sie zumindest im Zeitversatz kompensiert werden müssten. Weiterhin würde sich beim Vergleich im Zeitbereich eine mögliche Gesamtverstärkung ebenfalls als Störung auswirken, obwohl die Nutzsinalqualität davon nicht beeinflusst ist.

Aus diesem Grund erfolgt die Bewertung der Nutzsinalstörung durch Vergleich der gemittelten Frequenzgänge von Ausgangs- und Mikrofonsignal, die jeweils auf Einheitsvarianz normiert sind. Das auf die Standardabweichung σ normierte Ausgangssignal wird durch Kurzzeit-Fouriertransformation in den Frequenzbereich transformiert

$$\left(\frac{y_{i,s_i}}{\sigma_{y_{i,s_i}}}\right) \circ \bullet \tilde{Y}_{i,s_i}(f, n). \quad (6.6)$$

Für die Mikrofonsignale gilt äquivalent

$$\left(\frac{x_{i,s_i}}{\sigma_{x_{i,s_i}}}\right) \circ \bullet \tilde{X}_{i,s_i}(f, n). \quad (6.7)$$

Die Berechnung des mittleren Frequenzgangs von Ausgangs- und Mikrofonsignal erfolgt durch Mittelung über alle betrachteten Frames N_t als

$$\bar{Y}_i(f) = \frac{1}{N_t} \sum_{n=1}^{N_t} |\tilde{Y}_{i,s_i}(f, n)| \quad (6.8)$$

bzw. als

$$\bar{X}_i(f) = \frac{1}{N_t} \sum_{n=1}^{N_t} |\tilde{X}_{i,s_i}(f, n)|. \quad (6.9)$$

Die Unterschiede der mittleren Frequenzgänge lassen sich für alle Frequenzbänder unter Angabe der entsprechenden dB-Werte ermitteln. Für das Aufstellen eines geeigneten Bewertungsmaßes für die Nutzsinalstörung werden die einzelnen Abweichungen über die betrachteten Frequenzbänder gemittelt. Vor dem Hintergrund telefonbasierter Sprachanwendungen, die den Großteil elektronischer Kommunikationssysteme ausmachen, ist die Mittelung auf den relevanten Frequenzbereich von $f = 300$ Hz bis $f = 3400$ Hz begrenzt. Mit N_f als Anzahl der betrachteten Frequenzbänder ergibt sich das Maß für die Nutzsinalstörung als

$$\Delta\mathcal{S}^{(i)} = \frac{1}{N_f} \sum_{f=f_{300}}^{f_{3400}} \left(20 \log_{10} \frac{\bar{Y}_i(f)}{\bar{X}_i(f)} \right). \quad (6.10)$$

Die Nutzsinalverzerrung des frequenzvarianten Nullbeamforming weist, wenn nicht zusätzlich die Beträge der komplexen Faktoren optimiert werden, eine Hochpaßcharakteristik auf (vgl. Abschnitt 5.2.5.3). Diese Hochpaßcharakteristik äußert sich als konstanter Offset bei der Berechnung der Nutzsinalverzerrung nach Gleichung (6.10). Aus diesem Grund wird als zusätzliches Kriterium die Abweichung der Ausgangssignale von präemphasefilterten Originalsignalen bewertet. Das Präemphasefilter ist als Vorverarbeitung im Bereich der Spracherkennung weit verbreitet (z.B. [92]) und weist in diesem Fall die Koeffizienten

$$h_{pre} = [-1 \quad 0.97] \quad (6.11)$$

auf. Die gefilterten Mikrofonsignale

$$x_i^{(pre)} = h_{pre} * x_{i,s_i} \quad (6.12)$$

bilden die Grundlage für die Berechnung des gemittelten Nutzsinalfrequenzgangs

$$\bar{X}_i^{(pre)}(f) = \frac{1}{N_t} \sum_{n=1}^{N_t} |\tilde{X}_{i,s_i}^{(pre)}(f, n)|. \quad (6.13)$$

Damit läßt sich das zusätzliche Abstandsmaß in äquivalenter Form zu (6.10) angeben als

$$\Delta\mathcal{S}_{pre}^{(i)} = \frac{1}{N_f} \sum_{f=f_{300}}^{f_{3400}} \left(20 \log_{10} \frac{\bar{Y}_i(f)}{\bar{X}_i^{(pre)}(f)} \right). \quad (6.14)$$

Für die beiden Maße $\Delta\mathcal{S}^{(i)}$ und $\Delta\mathcal{S}_{pre}^{(i)}$ gilt, daß sie die Abweichung der Ausgangssignale von den Mikrofonsignalen widerspiegeln, aber keine Aussage über die Abweichung vom Quellsignal treffen. Dies korrespondiert mit der Problemstellung der blinden Quellentrennung, deren Kriterium der statistischen Unabhängigkeit keine Informationen über die Filterung eines Signals liefert.

6.2 Datensätze und Meßbedingungen

Alle der für die Auswertung verwendeten Aufnahmen stammen aus dem Datensatz der *TIDIGITS* [45], einer umfangreichen Sammlung von Ziffern in englischer Sprache, die von verschiedenen Sprechern wiedergegeben werden. Es handelt sich um rauscharme einkanalige Aufnahmen mit einer Abtastrate von 20 kHz und einer Auflösung von 16 Bit. Für die Auswertungen wurden Datensätze der Sprecher **Ar** (männlich, normale Stimmlage), **Fm** (männlich, hohe Stimmlage), **Ga** (männlich, tiefe Stimmlage) und **Ii** (weiblich) mit einer Länge von jeweils 10 Sekunden benutzt.

6.2.1 Erzeugung der synthetischen Mischungen

Für die synthetischen Mischungen wurden die Originalsignale auf eine Abtastrate von 11025 Hz reduziert, um sie besser mit den realen Raumaufnahmen vergleichen zu können. Die Mischung erfolgte mit künstlichen Impulsantworten, die mit Hilfe des Programms *room* [103] berechnet wurden. Als Simulationsgrundlage diente der bereits in Abschnitt 5.2.5.2 definierte Raum mit einem Volumen von 50 m³, bei gleicher Quellen- und Sensorpositionierung.

Für die Mischung wurden die Quellsignale mit den vier berechneten Impulsantworten zwischen jeweiliger Quelle und Mikrofon gefiltert und anschließend die gemeinsamen Anteile eines Sensors aufsummiert.

6.2.2 Aufnahme realer Mischungen

Um die Leistungsfähigkeit des Algorithmus in realen Umgebungen zu beurteilen, wurden Datensätze sowohl in einem Laborraum als auch in einem Automobil aufgezeichnet.

6.2.2.1 Aufnahmen im Laborraum

Die Dimension des Laborraumes beträgt ungefähr sechs mal sechs Meter bei einer Deckenhöhe von etwa 3.5 Metern, woraus sich ein Raumvolumen von über 100 m³ ergibt. Aufgrund der vielen glatten Flächen, wie Fenster und Boden, kann von einer Nachhallzeit von mehreren 100 ms ausgegangen werden.

Für die Aufnahmen wurden Meßmikrofone des Typs Behringer ECM 8000 mit Tascam MA-8-Vorverstärker sowie ein Harddisk-Recorder Alesis HD24 verwendet. Abbildung 6.1

zeigt den prinzipiellen Aufbau der Messungen mit einem Mikrofonabstand von 2.1 cm und den in einem Meter Entfernung platzierten Lautsprechern.

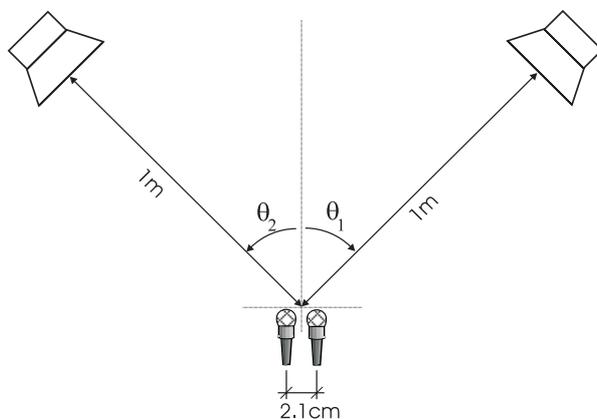


Abbildung 6.1: Meßanordnung im Laborraum

Die beiden sich durch Variation des Winkels θ_1 ergebenden Konfigurationen A und B sind in Tabelle 6.1 festgehalten. Zum Zweck der Auswertung wurden neben dem Signal mit beiden aktiven Sprechern auch die Einzelsignale der Sprecher aufgezeichnet.

Tabelle 6.1: Übersicht über die Konfigurationen der Messungen im Laborraum

| Konfiguration | θ_1 | θ_2 | Aufnahmen |
|---------------|-------------|------------|--|
| A | -45° | 25° | Sprecher 1, Sprecher 2, beide Sprecher |
| B | -10° | 25° | Sprecher 1, Sprecher 2, beide Sprecher |

6.2.2.2 Aufnahmen im Automobil

Die Aufnahmen im Automobil sind in Zusammenarbeit mit DaimlerChrysler entstanden. Die Aufzeichnung erfolgte in einem Mercedes S 320 mit omnidirektionalen Sennheiser-Mikrofonen ME 102 und einem 16-Kanal-Recorder bei einer Abtastrate von 12 kHz und einer Auflösung von 16 Bit. Die Sprachsignale wurden dabei mit Kunstköpfen von HEAD Acoustics abgespielt.

Abbildung 6.2 zeigt die möglichen Sprecherpositionen im Automobil. Dabei waren die Mikrofone über den einzelnen Sprechern sowie als lineares äquidistantes Array mit einem Mikrofonabstand von $d = 4$ cm in der Nähe des inneren Rückspiegels positioniert.

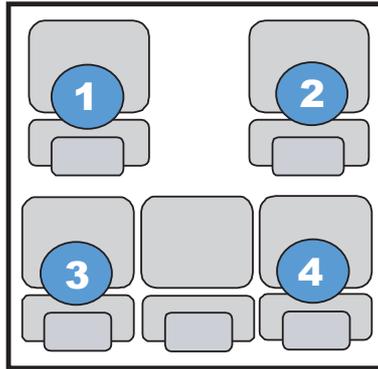


Abbildung 6.2: Positionen der Sprecher im Auto

Die Auswertung der Daten erfolgt für unterschiedliche Sprecherkombinationen und Aufnahmebedingungen. Tabelle 6.2 faßt die betrachteten Konfigurationen zusammen.

Tabelle 6.2: Konfigurationen der Messungen im Auto

| Konfiguration | Sprecher 1 und Position | Sprecher 2 und Position |
|---------------|-------------------------|-------------------------|
| Auto A | Ga auf 1 | Ii auf 2 |
| Auto B | Fm auf 3 | Ii auf 2 |
| Auto C | Ga auf 1 | Ar auf 4 |

6.3 Ergebnisse der synthetischen Mischungen

Nur mit Hilfe synthetischer Mischungen läßt sich die Abhängigkeit der Störunterdrückung von der Nachhallzeit des Raumes auswerten, sofern kein spezieller akustischer Raum mit veränderbarer Nachhallzeit zur Verfügung steht. Bei der Simulation wurden die Reflexionskoeffizienten des Raumes im Bereich von $\rho = 0.1$ bis $\rho = 0.9$ variiert. Die sich daraus ergebenden Nachhallzeiten lassen sich nach W. C. Sabine in Abhängigkeit vom Raumvolumen V und der Summe der Oberflächen des Raumes S berechnen als [57]

$$\tau_{60} = 0.163 \cdot \frac{V}{S \cdot (1 - \rho)}. \quad (6.15)$$

Die jeweiligen Nachhallzeiten sind in den Ergebnistabellen mit festgehalten.

Zusätzlich zu den Nachhallzeiten wurde sowohl der Mikrofonabstand als auch die Sprecherkombination verändert. Tabelle 6.5 bis 6.12 enthalten in detaillierter Form die erreichten SIR-Verbesserungen SIR_i^+ , die jeweilige Nutzsignalstörung $\Delta\mathcal{S}^{(i)}$, die jeweilige Nutzsignalstörung in Bezug auf präemphasegefilterte Mikrofonsignale $\Delta\mathcal{S}_{pre}^{(i)}$, sowie die benötigte Rechenzeit t_{calc} , jeweils für beide Kanäle. Die nachfolgende Tabelle gibt eine Übersicht über die dabei simulierten Sprecherkombinationen und die Mikrofonabstände.

Tabelle 6.3: Übersicht über die Ergebnistabellen der synthetischen Mischung

| Tabelle Nr. | Verfahren | Sprecherkombination | Mikrofonabstand |
|-------------|-----------|---------------------|-----------------|
| 6.5 | NBF | li / Fm | 2.5 cm |
| 6.6 | NBFb | li / Fm | 2.5 cm |
| 6.7 | NBF | li / Fm | 5 cm |
| 6.8 | NBFb | li / Fm | 5 cm |
| 6.9 | NBF | Ar / Ga | 2.5 cm |
| 6.10 | NBFb | Ar / Ga | 2.5 cm |
| 6.11 | NBF | Ar / Ga | 5 cm |
| 6.12 | NBFb | Ar / Ga | 5 cm |

Die Simulationen wurden sowohl mit dem auf reinen Phasenfaktoren beruhenden frequenzvarianten Nullbeamforming als auch unter zusätzlicher Berücksichtigung der Beträge, wie in Abschnitt 5.2.5.5 erläutert, durchgeführt. Die Bezeichnung der Verfahren ist in Tabelle 6.4 zusammengefasst.

Tabelle 6.4: Kurzbezeichnung der bei synthetischen Mischungen verwendeten Verfahren

| Bezeichnung | Verfahren |
|-------------|--|
| NBF | frequenzvariantes Nullbeamforming |
| NBFb | frequenzvariantes Nullbeamforming mit Berücksichtigung der Beträge |

6.3.1 Ergebnistabellen der synthetischen Mischungen

Tabelle 6.5: Ergebnisse für **NBF**, Sprecherkombination Ii/Fm, d=2.5cm in Abhängigkeit von der Nachhallzeit

| ρ | τ_{60} [ms] | SIR_1^+ [dB] | SIR_2^+ [dB] | $\Delta S^{(1)}$ [dB] | $\Delta S^{(2)}$ [dB] | $\Delta S_{pre}^{(1)}$ [dB] | $\Delta S_{pre}^{(2)}$ [dB] | t_{calc} [s] |
|--------|---------------------|-------------------|-------------------|--------------------------|--------------------------|--------------------------------|--------------------------------|-------------------|
| 0.1 | 107 | 26.20 | 22.38 | 5.70 | 5.96 | 0.25 | 0.26 | 0.92 |
| 0.2 | 120 | 23.52 | 19.90 | 5.67 | 6.00 | 0.41 | 0.38 | 0.94 |
| 0.3 | 137 | 20.54 | 17.02 | 5.65 | 6.04 | 0.59 | 0.51 | 0.91 |
| 0.4 | 160 | 17.43 | 13.88 | 5.64 | 6.06 | 0.76 | 0.66 | 0.94 |
| 0.5 | 192 | 14.07 | 10.49 | 5.59 | 6.05 | 1.01 | 0.81 | 1.00 |
| 0.6 | 240 | 10.83 | 7.76 | 5.54 | 5.95 | 1.21 | 1.07 | 1.02 |
| 0.7 | 320 | 7.82 | 4.98 | 5.44 | 5.67 | 1.49 | 1.23 | 0.98 |
| 0.8 | 479 | 4.21 | 2.36 | 5.02 | 5.03 | 1.70 | 1.67 | 1.05 |
| 0.9 | 959 | 1.87 | -0.44 | 4.30 | 4.24 | 1.87 | 2.08 | 1.09 |

Tabelle 6.6: Ergebnisse für **NBFb**, Sprecherkombination Ii/Fm, d=2.5cm in Abhängigkeit von der Nachhallzeit

| ρ | τ_{60} [ms] | SIR_1^+ [dB] | SIR_2^+ [dB] | $\Delta S^{(1)}$ [dB] | $\Delta S^{(2)}$ [dB] | $\Delta S_{pre}^{(1)}$ [dB] | $\Delta S_{pre}^{(2)}$ [dB] | t_{calc} [s] |
|--------|---------------------|-------------------|-------------------|--------------------------|--------------------------|--------------------------------|--------------------------------|-------------------|
| 0.1 | 107 | 21.48 | 17.63 | 0.23 | 0.14 | 5.55 | 5.99 | 0.91 |
| 0.2 | 120 | 21.89 | 18.06 | 0.23 | 0.20 | 5.55 | 6.03 | 0.98 |
| 0.3 | 137 | 22.08 | 18.16 | 0.25 | 0.28 | 5.55 | 6.05 | 1.03 |
| 0.4 | 160 | 21.46 | 16.43 | 0.29 | 0.37 | 5.50 | 6.08 | 1.09 |
| 0.5 | 192 | 16.23 | 14.01 | 0.32 | 0.56 | 5.52 | 6.16 | 1.19 |
| 0.6 | 240 | 12.24 | 11.34 | 0.58 | 0.86 | 5.56 | 6.20 | 1.23 |
| 0.7 | 320 | 8.69 | 6.09 | 1.15 | 1.05 | 5.47 | 6.16 | 1.22 |
| 0.8 | 479 | 3.06 | 3.54 | 1.44 | 2.03 | 5.39 | 6.66 | 1.25 |
| 0.9 | 959 | 0.62 | 0.01 | 2.38 | 2.04 | 6.63 | 4.72 | 1.16 |

Tabelle 6.7: Ergebnisse für **NBF**, Sprecherkombination Ii/Fm, d=5cm in Abhängigkeit von der Nachhallzeit

| ρ | τ_{60} [ms] | SIR_1^+ [dB] | SIR_2^+ [dB] | $\Delta\mathcal{S}^{(1)}$ [dB] | $\Delta\mathcal{S}^{(2)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(1)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(2)}$ [dB] | t_{calc} [s] |
|--------|---------------------|-------------------|-------------------|-----------------------------------|-----------------------------------|---|---|-------------------|
| 0.1 | 107 | 27.11 | 23.37 | 5.13 | 4.83 | 1.80 | 1.65 | 0.82 |
| 0.2 | 120 | 23.15 | 20.05 | 5.13 | 4.89 | 1.80 | 1.64 | 0.83 |
| 0.3 | 137 | 19.68 | 17.29 | 5.13 | 4.95 | 1.79 | 1.63 | 0.84 |
| 0.4 | 160 | 16.48 | 14.11 | 5.13 | 5.01 | 1.74 | 1.56 | 0.86 |
| 0.5 | 192 | 13.23 | 11.39 | 5.12 | 5.07 | 1.69 | 1.51 | 0.89 |
| 0.6 | 240 | 9.86 | 8.59 | 5.08 | 5.05 | 1.64 | 1.63 | 0.92 |
| 0.7 | 320 | 6.38 | 6.16 | 4.94 | 4.92 | 1.69 | 1.65 | 0.92 |
| 0.8 | 479 | 3.27 | 3.45 | 4.76 | 4.45 | 1.81 | 1.94 | 0.97 |
| 0.9 | 959 | 0.90 | 0.84 | 4.18 | 3.91 | 2.00 | 2.18 | 0.95 |

Tabelle 6.8: Ergebnisse für **NBFb**, Sprecherkombination Ii/Fm, d=5cm in Abhängigkeit von der Nachhallzeit

| ρ | τ_{60} [ms] | SIR_1^+ [dB] | SIR_2^+ [dB] | $\Delta\mathcal{S}^{(1)}$ [dB] | $\Delta\mathcal{S}^{(2)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(1)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(2)}$ [dB] | t_{calc} [s] |
|--------|---------------------|-------------------|-------------------|-----------------------------------|-----------------------------------|---|---|-------------------|
| 0.1 | 107 | 20.56 | 16.64 | 0.20 | 0.21 | 5.59 | 6.01 | 1.11 |
| 0.2 | 120 | 18.85 | 15.91 | 0.22 | 0.31 | 5.58 | 6.03 | 1.11 |
| 0.3 | 137 | 21.54 | 16.87 | 0.26 | 0.44 | 5.59 | 6.06 | 1.28 |
| 0.4 | 160 | 20.93 | 16.30 | 0.32 | 0.59 | 5.56 | 6.08 | 1.31 |
| 0.5 | 192 | 15.85 | 13.64 | 0.55 | 0.97 | 5.44 | 6.30 | 1.41 |
| 0.6 | 240 | 12.50 | 11.36 | 0.78 | 1.13 | 5.78 | 6.20 | 1.55 |
| 0.7 | 320 | 7.67 | 6.30 | 1.16 | 1.57 | 5.58 | 6.53 | 1.56 |
| 0.8 | 479 | 4.70 | 4.37 | 1.50 | 1.97 | 6.00 | 6.90 | 1.58 |
| 0.9 | 959 | 1.65 | 1.22 | 2.31 | 2.04 | 6.29 | 6.50 | 1.52 |

Tabelle 6.9: Ergebnisse für **NBF**, Sprecherkombination Ar/Ga, d=2.5cm in Abhängigkeit von der Nachhallzeit

| ρ | τ_{60} [ms] | SIR_1^+ [dB] | SIR_2^+ [dB] | $\Delta\mathcal{S}^{(1)}$ [dB] | $\Delta\mathcal{S}^{(2)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(1)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(2)}$ [dB] | t_{calc} [s] |
|--------|---------------------|-------------------|-------------------|-----------------------------------|-----------------------------------|---|---|-------------------|
| 0.1 | 107 | 21.11 | 19.64 | 6.44 | 7.42 | 0.29 | 0.27 | 0.80 |
| 0.2 | 120 | 20.08 | 17.91 | 6.48 | 7.40 | 0.41 | 0.40 | 0.83 |
| 0.3 | 137 | 18.54 | 15.67 | 6.53 | 7.40 | 0.56 | 0.57 | 0.86 |
| 0.4 | 160 | 16.31 | 13.31 | 6.56 | 7.40 | 0.73 | 0.75 | 0.86 |
| 0.5 | 192 | 13.64 | 10.51 | 6.54 | 7.36 | 0.92 | 0.92 | 0.89 |
| 0.6 | 240 | 10.73 | 7.64 | 6.49 | 7.28 | 1.13 | 1.17 | 0.95 |
| 0.7 | 320 | 7.93 | 4.82 | 6.21 | 7.05 | 1.51 | 1.32 | 0.94 |
| 0.8 | 479 | 4.97 | 2.37 | 5.47 | 6.56 | 1.85 | 1.66 | 0.95 |
| 0.9 | 959 | 2.71 | -0.34 | 5.01 | 5.60 | 2.05 | 1.84 | 1.05 |

Tabelle 6.10: Ergebnisse für **NBFb**, Sprecherkombination Ar/Ga, d=2.5cm in Abhängigkeit von der Nachhallzeit

| ρ | τ_{60} [ms] | SIR_1^+ [dB] | SIR_2^+ [dB] | $\Delta\mathcal{S}^{(1)}$ [dB] | $\Delta\mathcal{S}^{(2)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(1)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(2)}$ [dB] | t_{calc} [s] |
|--------|---------------------|-------------------|-------------------|-----------------------------------|-----------------------------------|---|---|-------------------|
| 0.1 | 107 | 19.11 | 21.26 | 0.12 | 0.23 | 6.25 | 7.14 | 1.08 |
| 0.2 | 120 | 19.82 | 21.09 | 0.13 | 0.28 | 6.27 | 7.14 | 1.11 |
| 0.3 | 137 | 20.21 | 20.63 | 0.14 | 0.36 | 6.31 | 7.13 | 1.14 |
| 0.4 | 160 | 19.62 | 17.69 | 0.17 | 0.44 | 6.35 | 7.12 | 1.22 |
| 0.5 | 192 | 14.43 | 12.07 | 2.48 | 0.87 | 8.71 | 6.64 | 0.89 |
| 0.6 | 240 | 12.37 | 8.33 | 0.63 | 0.81 | 6.39 | 7.01 | 1.22 |
| 0.7 | 320 | 8.11 | 9.32 | 1.17 | 1.00 | 7.21 | 6.95 | 1.23 |
| 0.8 | 479 | 0.24 | 0.12 | 0.79 | 0.62 | 5.26 | 7.00 | 0.81 |
| 0.9 | 959 | 0.37 | 0.21 | 1.96 | 2.14 | 3.64 | 7.67 | 1.08 |

Tabelle 6.11: Ergebnisse für **NBF**, Sprecherkombination Ar/Ga, d=5cm in Abhängigkeit von der Nachhallzeit

| ρ | τ_{60} [ms] | SIR_1^+ [dB] | SIR_2^+ [dB] | $\Delta\mathcal{S}^{(1)}$ [dB] | $\Delta\mathcal{S}^{(2)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(1)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(2)}$ [dB] | t_{calc} [s] |
|--------|---------------------|-------------------|-------------------|-----------------------------------|-----------------------------------|---|---|-------------------|
| 0.1 | 107 | 23.74 | 23.56 | 5.81 | 6.89 | 1.62 | 1.60 | 0.86 |
| 0.2 | 120 | 21.59 | 19.99 | 5.86 | 6.89 | 1.60 | 1.60 | 0.84 |
| 0.3 | 137 | 19.00 | 16.51 | 5.93 | 6.91 | 1.58 | 1.58 | 0.89 |
| 0.4 | 160 | 16.01 | 13.30 | 6.03 | 6.92 | 1.53 | 1.54 | 0.94 |
| 0.5 | 192 | 13.21 | 10.52 | 6.09 | 6.96 | 1.46 | 1.48 | 0.92 |
| 0.6 | 240 | 10.34 | 7.88 | 6.11 | 6.98 | 1.42 | 1.51 | 0.99 |
| 0.7 | 320 | 7.64 | 5.21 | 5.93 | 6.87 | 1.61 | 1.56 | 0.97 |
| 0.8 | 479 | 4.81 | 2.46 | 5.11 | 6.51 | 2.00 | 1.65 | 0.95 |
| 0.9 | 959 | 2.73 | -0.31 | 4.79 | 5.60 | 2.06 | 1.92 | 1.00 |

Tabelle 6.12: Ergebnisse für **NBFb**, Sprecherkombination Ar/Ga, d=5cm in Abhängigkeit von der Nachhallzeit

| ρ | τ_{60} [ms] | SIR_1^+ [dB] | SIR_2^+ [dB] | $\Delta\mathcal{S}^{(1)}$ [dB] | $\Delta\mathcal{S}^{(2)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(1)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(2)}$ [dB] | t_{calc} [s] |
|--------|---------------------|-------------------|-------------------|-----------------------------------|-----------------------------------|---|---|-------------------|
| 0.1 | 107 | 18.80 | 19.51 | 0.14 | 0.30 | 6.35 | 7.28 | 0.88 |
| 0.2 | 120 | 18.07 | 17.53 | 0.15 | 0.37 | 6.36 | 7.27 | 0.95 |
| 0.3 | 137 | 19.94 | 19.80 | 0.15 | 0.47 | 6.41 | 7.29 | 1.05 |
| 0.4 | 160 | 17.75 | 17.06 | 0.18 | 0.59 | 6.46 | 7.28 | 1.11 |
| 0.5 | 192 | 14.62 | 8.36 | 0.64 | 0.75 | 6.18 | 7.27 | 1.17 |
| 0.6 | 240 | 12.25 | 13.12 | 1.61 | 1.23 | 7.95 | 6.89 | 1.39 |
| 0.7 | 320 | 7.23 | 3.85 | 2.15 | 1.23 | 6.02 | 7.21 | 1.38 |
| 0.8 | 479 | 1.85 | 0.58 | 5.41 | 1.25 | 3.86 | 7.85 | 1.38 |
| 0.9 | 959 | -0.04 | -0.51 | 2.35 | 1.46 | 6.24 | 6.26 | 1.33 |

6.4 Ergebnisse der realen Aufnahmen

Neben dem in dieser Arbeit entwickelten Verfahren wurden für die Trennung der realen Raumaufnahmen als Referenz weitere Verfahren zur blinden Quellentrennung konvolutiver Mischungen benutzt. Es handelt sich zum einen um ein Frequenzbereichs-Verfahren, bei dem in jedem Band der Algorithmus JADE [24] für die Bestimmung der unabhängigen Komponenten mit einer anschließenden Permutationskorrektur entsprechend den Nullstellen im Beampattern genutzt wird.

Bei dem zweiten Verfahren handelt es sich um eine Implementierung des Algorithmus nach Parra [74], die unter [11] verfügbar ist. Dabei wurden die nötigen Parameter mit den folgenden Standardwerten initialisiert: Länge der Zeitfilter = 128, FFT-Länge = 512, Anzahl der zu diagonalisierenden Matrizen = 5, Anzahl der Intervalle, die für die Bestimmung einer Kreuzleistungsdichte-Matrix genutzt werden = 7.

Die nachfolgende Tabelle faßt die Bezeichnung der Verfahren zusammen.

Tabelle 6.13: Kurzbezeichnung der bei realen Aufnahmen verwendeten Verfahren

| Bezeichnung | Verfahren |
|-------------|--|
| NBF | frequenzvariantes Nullbeamforming |
| NBFb | frequenzvariantes Nullbeamforming mit Berücksichtigung der Beträge |
| JADE | JADE im Frequenzbereich mit beampatternbasierter Permutationskorrektur |
| PARRA | Verfahren nach Parra und Spence [74] |

In Tabelle 6.14 bis 6.18 sind die Ergebnisse des Vergleichs der einzelnen Verfahren dokumentiert. Die Bewertungskriterien sind dabei, identisch zur Auswertung der synthetischen Mischungen, die SIR-Verbesserung SIR_i^+ , die Nutzsignalstörung $\Delta\mathcal{S}^{(i)}$, die Nutzsignalstörung in Bezug auf präemphasegefilterte Mikrofonssignale $\Delta\mathcal{S}_{pre}^{(i)}$, sowie die benötigte Rechenzeit t_{calc} . Die jeweils besten Werte, d.h. das größte SIR, die kleinste Nutzsignalverzerrung, sowie die geringste Rechenzeit sind fett hervorgehoben.

Tabelle 6.14: Ergebnisse im Laborraum bei Konfiguration A

| Algorithmus | SIR_1^+ [dB] | SIR_2^+ [dB] | $\Delta\mathcal{S}^{(1)}$ [dB] | $\Delta\mathcal{S}^{(2)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(1)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(2)}$ [dB] | t_{calc} [s] |
|-------------|-------------------|-------------------|-----------------------------------|-----------------------------------|---|---|-------------------|
| NBF | 6.08 | 3.35 | 5.06 | 6.09 | 1.19 | 1.53 | 1.31 |
| NBFb | 2.32 | 5.34 | 0.49 | 0.70 | 6.31 | 5.26 | 1.39 |
| Jade | 1.67 | 4.38 | 1.37 | 0.42 | 7.08 | 5.56 | 3.94 |
| Parra | 0.69 | 3.66 | 1.82 | 4.49 | 4.65 | 2.11 | 5.40 |

Tabelle 6.15: Ergebnisse im Laborraum bei Konfiguration B

| Algorithmus | SIR_1^+ [dB] | SIR_2^+ [dB] | $\Delta\mathcal{S}^{(1)}$ [dB] | $\Delta\mathcal{S}^{(2)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(1)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(2)}$ [dB] | t_{calc} [s] |
|-------------|-------------------|-------------------|-----------------------------------|-----------------------------------|---|---|-------------------|
| NBF | 3.88 | 4.11 | 5.43 | 4.91 | 1.58 | 1.07 | 1.48 |
| NBFb | 0.83 | 0.99 | 3.65 | 3.05 | 8.42 | 5.88 | 1.61 |
| Jade | -0.97 | 5.50 | 0.95 | 1.50 | 6.69 | 6.54 | 4.42 |
| Parra | 1.80 | 0.63 | 2.47 | 5.67 | 3.55 | 1.82 | 7.40 |

Tabelle 6.16: Ergebnisse im Kfz bei Konfiguration A

| Algorithmus | SIR_1^+ [dB] | SIR_2^+ [dB] | $\Delta\mathcal{S}^{(1)}$ [dB] | $\Delta\mathcal{S}^{(2)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(1)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(2)}$ [dB] | t_{calc} [s] |
|-------------|-------------------|-------------------|-----------------------------------|-----------------------------------|---|---|-------------------|
| NBF | 10.04 | 10.28 | 3.57 | 6.17 | 1.68 | 2.27 | 1.77 |
| NBFb | 4.79 | 10.42 | 2.11 | 2.68 | 3.63 | 5.88 | 1.77 |
| Jade | 11.47 | 15.60 | 0.67 | 1.36 | 5.00 | 8.57 | 3.91 |
| Parra | 2.84 | 0.97 | 2.06 | 4.30 | 3.13 | 3.70 | 5.19 |

6.5 Diskussion

Die Ergebnisse der synthetischen Mischungen bestätigen die theoretischen Überlegungen aus Kapitel 5. Bei längerer Nachhallzeit geht die Eindeutigkeit der Einfallstrichtung verloren, und es dauert länger, die Ausnullungsrichtungen zu bestimmen. Abbildung 6.3 verdeutlicht diesen Zusammenhang beispielhaft anhand der benötigten Rechenzeit eines Durchgangs für die beiden Verfahren NBF und NBFb.

Tabelle 6.17: Ergebnisse im Kfz bei Konfiguration B

| Algorithmus | SIR_1^+ [dB] | SIR_2^+ [dB] | $\Delta\mathcal{S}^{(1)}$ [dB] | $\Delta\mathcal{S}^{(2)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(1)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(2)}$ [dB] | t_{calc} [s] |
|-------------|-------------------|-------------------|-----------------------------------|-----------------------------------|---|---|-------------------|
| NBF | 5.69 | 1.76 | 6.40 | 6.14 | 2.54 | 2.36 | 1.50 |
| NBFb | 3.71 | 2.66 | 2.82 | 2.73 | 5.46 | 5.81 | 2.03 |
| Jade | 4.76 | 5.49 | 2.36 | 2.38 | 7.26 | 9.41 | 4.20 |
| Parra | 2.44 | -0.62 | 4.47 | 6.19 | 2.80 | 2.69 | 6.12 |

Tabelle 6.18: Ergebnisse im Kfz bei Konfiguration C

| Algorithmus | SIR_1^+ [dB] | SIR_2^+ [dB] | $\Delta\mathcal{S}^{(1)}$ [dB] | $\Delta\mathcal{S}^{(2)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(1)}$ [dB] | $\Delta\mathcal{S}_{pre}^{(2)}$ [dB] | t_{calc} [s] |
|-------------|-------------------|-------------------|-----------------------------------|-----------------------------------|---|---|-------------------|
| NBF | 5.05 | 12.12 | 3.96 | 5.55 | 1.53 | 1.75 | 1.66 |
| NBFb | 0.60 | 9.99 | 2.40 | 4.12 | 3.33 | 4.61 | 2.25 |
| Jade | 6.62 | 4.49 | 1.81 | 3.44 | 5.85 | 8.89 | 4.02 |
| Parra | 0.70 | 2.85 | 5.03 | 2.47 | 1.63 | 3.53 | 5.85 |

Weiterhin wird deutlich, daß auch die Trennungsleistung stark von der Nachhallzeit abhängt. Abbildung 6.4 auf der nächsten Seite zeigt das für beide Kanäle gemittelte SIR in Abhängigkeit von den simulierten Reflexionskoeffizienten und von der Nachhallzeit bei simulierten Mikrofonabständen von 25 bzw. 50 mm.

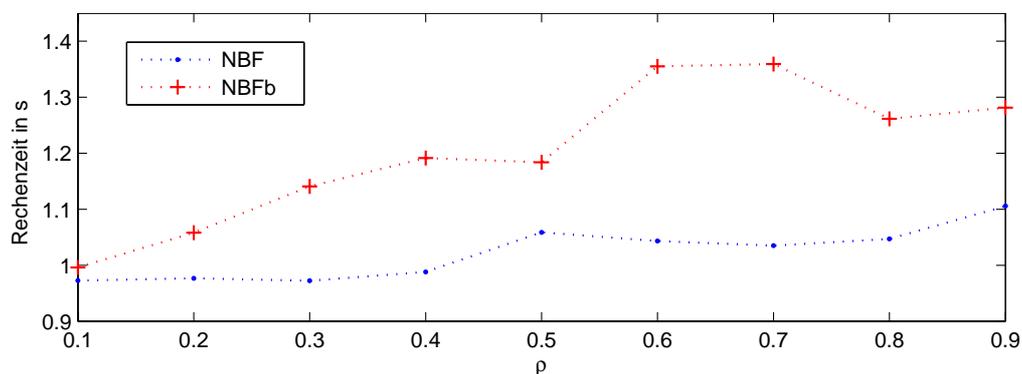


Abbildung 6.3: Beispiel für die Abhängigkeit der Rechenzeit von der Nachhallzeit bei synthetischen Mischungen

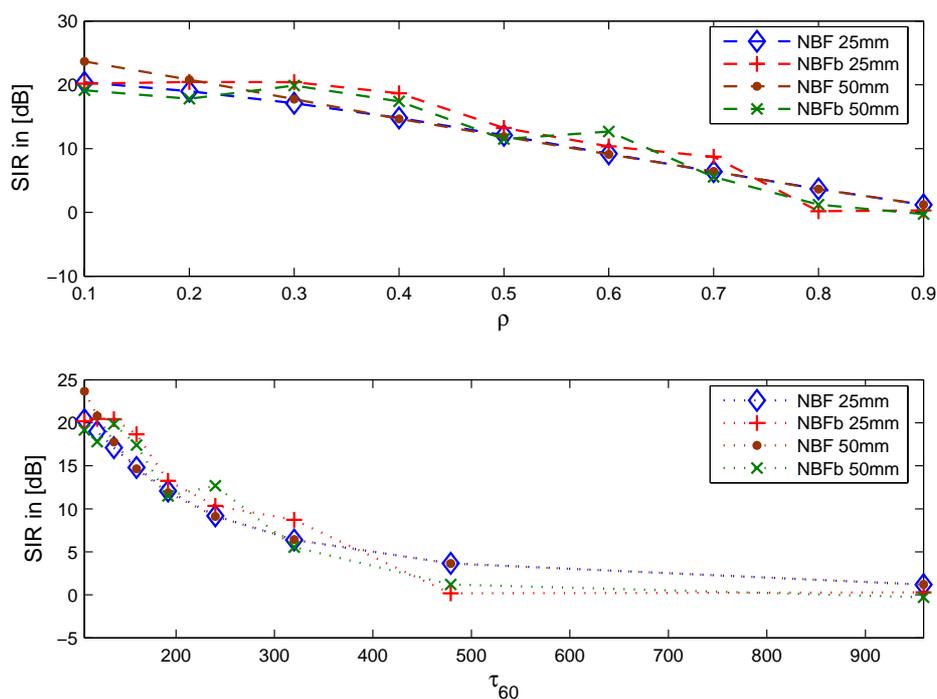


Abbildung 6.4: Abhängigkeit der Trennungsleistung von der Nachhallzeit

Die nachlassende Trennungsleistung kann durch eine Fehlanpassung des Modells an die realen Gegebenheiten erklärt werden. Dabei spielt die unzureichende Fensterlänge eine entscheidende Rolle, wenn man berücksichtigt, daß bei einer Abtastrate von 11 kHz die Fensterlänge von 512 Taps einem Zeitabschnitt von 46.4 ms entspricht. Nur die innerhalb dieses Zeitfensters liegenden Anteile der Raumimpulsantwort lassen sich nach entsprechender Kurzzeit-Fouriertransformation erfassen (vgl. Abschnitt 2.2). Da Nachhallzeiten einen weitaus größeren Zeitraum in Anspruch nehmen können, muß die Trennungsleistung für größere Nachhallzeiten zwangsläufig schlechter werden. In Abbildung 6.5 ist das mittlere SIR bei Sprecherkombination Fm / Ii und Mikrofonabstand $d = 5$ cm für verschiedene FFT-Längen von 32 bis 2048 und die Reflexionskoeffizienten $\rho=0.3$ und $\rho=0.5$ dargestellt.

Mit steigender FFT-Länge nimmt die Trennungsqualität zu, so daß mit dem Verfahren NBFb bei einer Nachhallzeit von 192 ms ($\rho=0.5$) das Optimum erst bei einer FFT-Länge von 1024 Werten erreicht wird, wohingegen bei einer Nachhallzeit von 137 ms ($\rho=0.3$) nur 256 Taps nötig sind.

Das Verfahren NBF zeigt ebenfalls diese Form der Abhängigkeit, bleibt für längere Nachhallzeiten im Trennungsergebnis aber hinter der Variante mit zusätzlicher Betragsberück-

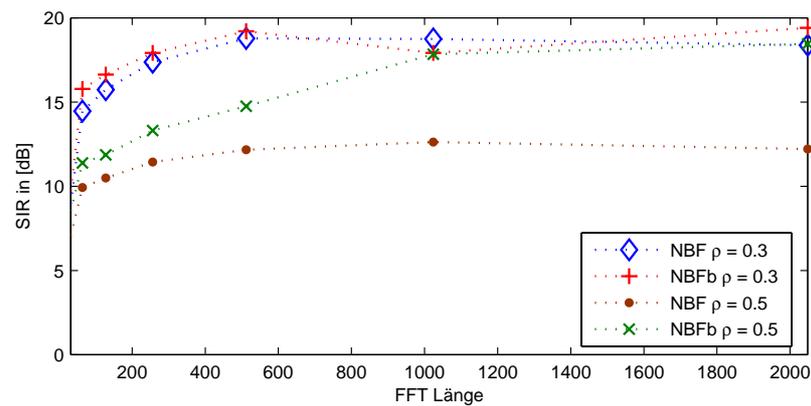


Abbildung 6.5: Abhängigkeit der Trennungsleistung von der FFT-Länge

sichtigung zurück. Dieses Verhalten ist mit der für größere Nachhallzeiten zunehmenden Variation der Beträge zu erklären, die durch die reine Phasenverschiebung nicht kompensiert werden können (vgl. Abschnitt 5.2.5.5).

Für sämtliche Frequenzbereichsverfahren ist jedoch zu beachten, daß die FFT-Länge nicht beliebig vergrößert werden kann. Zum einen erhöht sich die Rechenzeit mit der Anzahl der betrachteten Frequenzbänder und zum anderen muß man einen höheren Zeitversatz zwischen Eingangssignal und Ausgangssignal in Kauf nehmen, der einer realzeitfähigen Implementierung im Wege steht.

Bei Auswertung der Ergebnisse der realen Aufnahmen läßt sich festhalten, daß die beiden Verfahren des frequenzvarianten Nullbeamforming im Vergleich sehr gute Trennungsleistungen erzielen. Die Leistungsfähigkeit ist in etwa zum Verfahren mit dem JADE-Algorithmus vergleichbar. Sie sollte tatsächlich eher geringer sein, da im Vergleich zu JADE nicht alle Kreuzkumulanten ausgewertet werden und keine Dekorrelation der Daten durch Prewhitening erzwungen wird.

Dieser Nachteil kann im Hinblick auf die benötigte Rechenleistung in Kauf genommen werden. Obwohl bei zusätzlicher Berücksichtigung des Betrages die vorteilhafte Parametrisierung des phasenbasierten Verfahrens nicht genutzt werden kann, ist diese Variante im Vergleich nur ca. 30 % langsamer. Beide Algorithmen sind in etwa doppelt so schnell wie das JADE-basierte Verfahren und drei mal so schnell wie das Verfahren nach Parra.

Bei der Analyse der Werte der erreichten SIR-Verbesserung unter realen Bedingungen erweist sich das frequenzvariante Nullbeamforming in den verschiedenen Konfigurationen als sehr zuverlässig. Die mittlere SIR-Verbesserung beträgt bis zu 10 dB und übertrifft

in einigen Fällen die Ergebnisse des JADE-Algorithmus. Auffällig ist, daß die erreichte Dämpfung nicht wie bei den synthetischen Mischungen für beide Quellen in etwa gleich ist. Vielmehr unterscheiden sich die Übertragungsfunktionen unter realen Bedingungen in manchen Fällen so stark, daß nur für einen Kanal eine gute SIR-Verbesserung möglich ist (siehe Ergebnisse für NFBb in Tabelle 6.18).

Bei erfolgreicher Entmischung zeichnen sich alle Verfahren der blinden Quellentrennung durch eine sehr geringe Nutzsignalverzerrung aus. In dieser Hinsicht bietet das betragsbasierte Verfahren den Vorteil, die Signale nicht durch Filterung mit Hochpaßcharakteristik zu verändern. Dadurch bleibt die durchschnittliche Abweichung vom Nutzsignalspektrum gering.

Der Vergleich der mit dem phasenbasierten Verfahren berechneten Ausgangssignale mit den präemphasefilterten Mikrofonsignalen zeigt dagegen, daß die Veränderung des Nutzsinalsignals Hochpaßcharakter aufweist (siehe Abschnitt 5.2.5.3). Die Verständlichkeit der Sprache wird dadurch nicht verschlechtert, und die Weiterverarbeitung mit einem Spracherkennung, der präemphasefilterte Signale benötigt, ist möglich.

7 Zusammenfassung

In der vorliegenden Arbeit wurde das Verfahren des frequenzvarianten Nullbeamforming entwickelt, das als Kombination der Konzepte Beamforming und blinde Quellentrennung verstanden werden kann. Die Vorteile dieses Verfahrens sind der im Vergleich zu den herkömmlichen Methoden der blinden Quellentrennung konvolutiver Mischungen geringere Rechenaufwand und die im Vergleich zu adaptiven Beamformern geringe Nutzsignalverzerrung. Damit ist es gelungen, die Vorzüge dieser beiden Ansätze in einem robusten Verfahren zu vereinen, das für den Einsatz in realzeitfähigen Systemen geeignet ist.

Für die Herleitung des frequenzvarianten Nullbeamforming wurden die Prinzipien der mehrkanaligen Störgeräuschunterdrückung analysiert, um ein grundlegendes Verständnis für die Funktionsweise von Beamforming-Verfahren für Mikrofonarrays zu schaffen. Adaptive Beamformer weisen einige Analogien zu den Methoden der blinden Quellentrennung konvolutiver Mischungen auf. Sie kommen im Vergleich zu fest entworfenen Arrays mit deutlich weniger Mikrofonen für die Unterdrückung gerichteter Störsignale aus. Nachteilig ist jedoch, daß bei einer fehlerhaften Adaption, insbesondere bedingt durch unrealistische Vorgaben der Nutzsignalrichtung, das interessierende Sprachsignal verzerrt werden kann. Verfahren, die nicht nur die räumliche Charakteristik des Stör- sondern auch des Nutzsignals nachregeln können, also alle beteiligten Signale berücksichtigen, haben diesen Nachteil nicht.

Ein äußerst leistungsfähiges Kriterium zur Unterscheidung zwischen gerichteten Signalen ist durch den Grad der statistischen Unabhängigkeit gegeben, der auch in anderen Bereichen der Signalverarbeitung erfolgreich für die blinde Trennung von Signalgemischen eingesetzt wird. Durch das Kriterium der statistischen Unabhängigkeit lassen sich die intersensorischen Beziehungen der einzelnen Signalkomponenten aufdecken, die für die Rekonstruktion der Quellen erforderlich sind. Im Kontext von Mikrofonarrays können diese Parameter als räumliche Größen interpretiert werden, da die Dämpfung eines gerichteten Signals durch das Ausnullen der entsprechenden Einfallrichtung erreicht wird.

Dies ist die Grundlage für das in dieser Arbeit entwickelte Verfahren des frequenzvarianten Nullbeamforming, bei dem die auszunullenden Richtungen mit Hilfe von Statistik höherer Ordnung so bestimmt werden, daß die Ausgangssignale statistisch unabhängig voneinander sind. Durch Parametrisierung als Beamforming-Phasenfaktoren läßt sich nicht nur der Suchraum eingrenzen, sondern auch ein effizienter Algorithmus auf der Basis eines Gradientenabstiegs formulieren.

Die Wirksamkeit dieses Verfahrens konnte anhand praktischer Versuche verifiziert werden. Dabei wurde gezeigt, daß die erreichbare Störsignalunterdrückung von der akustischen Umgebung abhängig ist. Wenn die zwischen Quelle und Mikrofon meßbare Nachhallzeit in einer mit Hilfe des Modells erfaßbaren Größenordnung liegt, kann eine robuste Störsignalunterdrückung realisiert werden.

In realen Umgebungen hat man es zumeist mit einer Vielzahl von Reflexionen zu tun, die sich aufgrund der starken Abweichung vom Fernfeld-Beamforming-Modell verschlechternd auf die Eindeutigkeit der Einfallsrichtungen auswirken. Dieser Zusammenhang wurde anhand der Trennungsergebnisse bei Aufnahmen in realen Umgebungen bestätigt, und er eröffnet gleichzeitig die Möglichkeit, in zukünftigen Verfahren gezielt auf diese Problematik Einfluß zu nehmen. Nur so können die offenen Herausforderungen, wie z.B. bewegte Signalquellen und lange Nachhallzeiten, in Zukunft gelöst werden.

A Gradientenberechnung

Für die Berechnung des Gradienten der Kostenfunktion aus Gleichung (5.29) müssen die partiellen Ableitungen der Kreuzkumulante vierter Ordnung bezüglich φ_1 und φ_2 bestimmt werden. Durch Multiplikation des Gradienten mit dem Vorzeichen der Kreuzkumulante ist die betragsmäßige Minimierung der Kostenfunktion sichergestellt.

Die Ausgangssignale des frequenzvarianten Nullbeamforming ergeben sich in jedem Frequenzband zu

$$\begin{aligned} Y_1(\omega) &= -e_2(\varphi)X_1(\omega) + X_2(\omega) \\ Y_2(\omega) &= e_1(\varphi)X_1(\omega) - X_2(\omega). \end{aligned}$$

Nach Normierung der Ausgangssignale auf Einheitsvarianz läßt sich die in Gleichung (5.28) definierte Kreuzkumulante vierter Ordnung darstellen als

$$\widehat{\text{cum}}_{22}(\tilde{Y}_1, \tilde{Y}_2) = E[\tilde{Y}_1 \cdot \tilde{Y}_1^* \cdot \tilde{Y}_2 \cdot \tilde{Y}_2^*] - E[\tilde{Y}_1 \tilde{Y}_2^*] \cdot E[\tilde{Y}_2 \tilde{Y}_1^*] - E[\tilde{Y}_1 \tilde{Y}_2] \cdot E[\tilde{Y}_1^* \tilde{Y}_2^*].$$

Mit Hilfe der drei Terme \mathcal{K}_1 , \mathcal{K}_2 und \mathcal{K}_3 kann die Kreuzkumulante auch wie folgt formuliert werden:

$$\widehat{\text{cum}}_{22}(\tilde{Y}_1, \tilde{Y}_2) = \mathcal{K}_1(\omega) - \mathcal{K}_2(\omega) - \mathcal{K}_3(\omega). \quad (\text{A.1})$$

Dabei sind die drei Terme von der Form

$$\mathcal{K}_1(\omega) = E[|\tilde{Y}_1(\omega)|^2 \cdot |\tilde{Y}_2(\omega)|^2] = E[\tilde{Y}_1(\omega) \cdot \tilde{Y}_1^*(\omega) \cdot \tilde{Y}_2(\omega) \cdot \tilde{Y}_2^*(\omega)] \quad (\text{A.2})$$

$$\mathcal{K}_2(\omega) = \left| E[\tilde{Y}_1(\omega) \tilde{Y}_2^*(\omega)] \right|^2 = E[\tilde{Y}_1(\omega) \tilde{Y}_2^*(\omega)] \cdot E[\tilde{Y}_1^*(\omega) \tilde{Y}_2(\omega)] \quad (\text{A.3})$$

$$\mathcal{K}_3(\omega) = \left| E[\tilde{Y}_1(\omega) \tilde{Y}_2(\omega)] \right|^2 = E[\tilde{Y}_1(\omega) \tilde{Y}_2(\omega)] \cdot E[\tilde{Y}_1^*(\omega) \tilde{Y}_2^*(\omega)]. \quad (\text{A.4})$$

Sie lassen sich für die Berechnung des Gradienten getrennt voneinander auswerten. Dazu werden nachfolgend die partiellen Ableitungen von \mathcal{K}_1 , \mathcal{K}_2 und \mathcal{K}_3 nach den konjugiert komplexen Variablen e_1^* und e_2^* aufgestellt [18], wobei aus Gründen der Übersichtlichkeit die Abhängigkeit von ω nicht explizit mitgeführt wird.

Ableitung von \mathcal{K}_1 nach e_1^*

$$\begin{aligned}
\frac{\partial \mathcal{K}_1}{\partial e_1^*} &= \frac{\partial}{\partial e_1^*} \left(E[\tilde{Y}_1 \tilde{Y}_1^* \tilde{Y}_2 \tilde{Y}_2^*] \right) \\
&= \frac{\partial}{\partial e_1^*} \left(E \left[\frac{Y_1 Y_1^* Y_2 Y_2^*}{E[Y_1 Y_1^*] E[Y_2 Y_2^*]} \right] \right) \\
&= E \left[\frac{Y_1 Y_1^* Y_2}{E[Y_1 Y_1^*]} \cdot \frac{X_1^* \sigma_{Y_2}^2 - Y_2^* E[Y_2 X_1^*]}{E[Y_2 Y_2^*]^2} \right] \\
&= \frac{E[Y_1 Y_1^* Y_2 \cdot (X_1^* \sigma_{Y_2}^2 - Y_2^* E[Y_2 X_1^*])]}{\sigma_{Y_1}^2 \sigma_{Y_2}^4} \\
&= \frac{E[Y_1 Y_1^* Y_2 X_1^*] \sigma_{Y_2}^2 - E[Y_1 Y_1^* Y_2 Y_2^* E[Y_2 X_1^*]]}{\sigma_{Y_1}^2 \sigma_{Y_2}^4}
\end{aligned} \tag{A.5}$$

Ableitung von \mathcal{K}_1 nach e_2^*

$$\begin{aligned}
\frac{\partial \mathcal{K}_1}{\partial e_2^*} &= \frac{\partial}{\partial e_2^*} \left(E \left[\frac{Y_1 Y_1^* Y_2 Y_2^*}{E[Y_1 Y_1^*] E[Y_2 Y_2^*]} \right] \right) \\
&= E \left[\frac{Y_1 Y_2 Y_2^*}{E[Y_2 Y_2^*]} \cdot \frac{-X_1^* \sigma_{Y_1}^2 - Y_1^* E[Y_1 (-X_1^*)]}{E[Y_1 Y_1^*]^2} \right] \\
&= \frac{E[Y_1 Y_1^* Y_2 Y_2^* E[Y_1 X_1^*]] - E[Y_1 Y_2 Y_2^* X_1^*] \sigma_{Y_1}^2}{\sigma_{Y_1}^4 \sigma_{Y_2}^2}
\end{aligned} \tag{A.6}$$

Ableitung von \mathcal{K}_2 nach e_1^*

$$\begin{aligned}
\frac{\partial \mathcal{K}_2}{\partial e_1^*} &= \frac{\partial}{\partial e_1^*} \left(E[\tilde{Y}_1 \tilde{Y}_2^*] \cdot E[\tilde{Y}_2 \tilde{Y}_1^*] \right) \\
&= \frac{\partial}{\partial e_1^*} \left(\frac{E[Y_1 Y_2^*] E[Y_1^* Y_2]}{E[Y_1 Y_1^*] E[Y_2 Y_2^*]} \right) \\
&= \frac{E[Y_1 X_1^*] E[Y_1^* Y_2] \cdot E[Y_1 Y_1^*] E[Y_2 Y_2^*] - E[Y_1 Y_2^*] E[Y_1^* Y_2] \cdot E[Y_1 Y_1^*] E[Y_2 X_1^*]}{E[Y_1 Y_1^*]^2 \cdot E[Y_2 Y_2^*]^2} \\
&= \frac{E[Y_1 X_1^*] E[Y_1^* Y_2] \sigma_{Y_2}^2 - E[Y_1 Y_2^*] E[Y_1^* Y_2] E[Y_2 X_1^*]}{\sigma_{Y_1}^2 \sigma_{Y_2}^4}
\end{aligned} \tag{A.7}$$

Ableitung von \mathcal{K}_2 nach e_2^*

$$\begin{aligned}
\frac{\partial \mathcal{K}_2}{\partial e_2^*} &= \frac{\partial}{\partial e_2^*} \left(\frac{E[Y_1 Y_2^*] E[Y_1^* Y_2]}{E[Y_1 Y_1^*] E[Y_2 Y_2^*]} \right) \\
&= \frac{E[Y_1 Y_2^*] E[-X_1^* Y_2] \cdot E[Y_1 Y_1^*] E[Y_2 Y_2^*] - E[Y_1 Y_2^*] E[Y_1^* Y_2] \cdot E[Y_1(-X_1^*)] E[Y_2 Y_2^*]}{E[Y_1 Y_1^*]^2 \cdot E[Y_2 Y_2^*]^2} \\
&= \frac{E[Y_1 Y_2^*] E[Y_1^* Y_2] E[Y_1 X_1^*] - E[Y_1 Y_2^*] E[X_1^* Y_2] \sigma_{Y_1}^2}{\sigma_{Y_1}^4 \sigma_{Y_2}^2} \tag{A.8}
\end{aligned}$$

Ableitung von \mathcal{K}_3 nach e_1^*

$$\begin{aligned}
\frac{\partial \mathcal{K}_3}{\partial e_1^*} &= \frac{\partial}{\partial e_1^*} \left(E[\tilde{Y}_1 \tilde{Y}_2] \cdot E[\tilde{Y}_1^* \tilde{Y}_2^*] \right) \\
&= \frac{\partial}{\partial e_1^*} \left(\frac{E[Y_1 Y_2] E[Y_1^* Y_2^*]}{E[Y_1 Y_1^*] E[Y_2 Y_2^*]} \right) \\
&= \frac{E[Y_1 Y_2] E[Y_1^* X_1^*] \cdot E[Y_1 Y_1^*] E[Y_2 Y_2^*] - E[Y_1 Y_2] E[Y_1^* Y_2^*] \cdot E[Y_1 Y_1^*] E[Y_2 X_1^*]}{E[Y_1 Y_1^*]^2 E[Y_2 Y_2^*]^2} \\
&= \frac{E[Y_1 Y_2] E[Y_1^* X_1^*] \sigma_{Y_2}^2 - E[Y_1 Y_2] E[Y_1^* Y_2^*] E[Y_2 X_1^*]}{\sigma_{Y_1}^2 \sigma_{Y_2}^4} \tag{A.9}
\end{aligned}$$

Ableitung von \mathcal{K}_3 nach e_2^*

$$\begin{aligned}
\frac{\partial \mathcal{K}_3}{\partial e_2^*} &= \frac{\partial}{\partial e_2^*} \left(\frac{E[Y_1 Y_2] E[Y_1^* Y_2^*]}{E[Y_1 Y_1^*] E[Y_2 Y_2^*]} \right) \\
&= \frac{E[Y_1 Y_2] E[Y_2^*(-X_1^*)] \cdot E[Y_1 Y_1^*] E[Y_2 Y_2^*] - E[Y_1 Y_2] E[Y_1^* Y_2^*] \cdot E[Y_1(-X_1^*)] E[Y_2 Y_2^*]}{E[Y_1 Y_1^*]^2 E[Y_2 Y_2^*]^2} \\
&= \frac{E[Y_1 Y_2] E[Y_1^* Y_2^*] E[Y_1 X_1^*] - E[Y_1 Y_2] E[Y_2^* X_1^*] \sigma_{Y_1}^2}{\sigma_{Y_1}^4 \sigma_{Y_2}^2} \tag{A.10}
\end{aligned}$$

Mit diesen Ergebnissen lassen sich die partiellen Ableitungen der modifizierten Kreuzkumulante nach e_1^* und e_2^* wie folgt angeben

$$\begin{aligned} \frac{\partial \widehat{\text{cum}}_{22}(\tilde{Y}_1, \tilde{Y}_2)}{\partial e_1^*} &= \frac{1}{\sigma_{Y_1}^2 \sigma_{Y_2}^4} \cdot \left(E[Y_1 Y_1^* Y_2 X_1^*] \sigma_{Y_2}^2 - E[Y_1 Y_1^* Y_2 Y_2^* E[Y_2 X_1^*]] - E[Y_1 X_1^*] E[Y_1^* Y_2] \sigma_{Y_2}^2 \right. \\ &\quad \left. + \left| E[Y_1 Y_2^*] \right|^2 E[Y_2 X_1^*] - E[Y_1 Y_2] E[Y_1^* X_1^*] \sigma_{Y_2}^2 + \left| E[Y_1 Y_2] \right|^2 E[Y_2 X_1^*] \right) \end{aligned} \quad (\text{A.11})$$

und

$$\begin{aligned} \frac{\partial \widehat{\text{cum}}_{22}(\tilde{Y}_1, \tilde{Y}_2)}{\partial e_2^*} &= \frac{1}{\sigma_{Y_1}^4 \sigma_{Y_2}^2} \cdot \left(E[Y_1 Y_1^* Y_2 Y_2^* E[Y_1 X_1^*]] - E[Y_1 Y_2 Y_2^* X_1^*] \sigma_{Y_1}^2 - \left| E[Y_1 Y_2^*] \right|^2 E[Y_1 X_1^*] \right. \\ &\quad \left. + E[Y_1 Y_2^*] E[X_1^* Y_2] \sigma_{Y_1}^2 - \left| E[Y_1 Y_2] \right|^2 E[Y_1 X_1^*] + E[Y_1 Y_2] E[Y_2^* X_1^*] \sigma_{Y_1}^2 \right). \end{aligned} \quad (\text{A.12})$$

B Reduktion des Rechenaufwandes

Für eine effiziente Implementierung des Algorithmus ist es notwendig, die rechenaufwendigen Erwartungswertbildungen zwischen den Ausgangssignalen auf Erwartungswertbildungen zwischen Meßsignalen zurückzuführen. Dafür sind insgesamt 12 verschiedene Erwartungswerte zu berechnen, die nachfolgend aus Gründen der Übersichtlichkeit mit ν_1 bis ν_{12} bezeichnet werden und wie folgt definiert sind:

$$\begin{aligned}
 \nu_1 &= E[X_1^2] & \nu_2 &= E[X_2^2] \\
 \nu_3 &= E[|X_1|^2] & \nu_4 &= E[|X_2|^2] \\
 \nu_5 &= E[X_1 X_2] & \nu_6 &= E[X_1 X_2^*] \\
 \nu_7 &= E[X_1^2 X_2^{*2}] & \nu_8 &= E[|X_1|^2 |X_2|^2] \\
 \nu_9 &= E[X_1 X_2^* |X_1|^2] & \nu_{10} &= E[X_1 X_2^* |X_2|^2] \\
 \nu_{11} &= E[|X_1|^4] & \nu_{12} &= E[|X_2|^4].
 \end{aligned}$$

B.1 Aufwandsreduzierte Kostenfunktion

Bei der Herleitung der aufwandsreduzierten Kostenfunktion müssen die in den Gleichungen (A.2) bis (A.4) definierten Terme \mathcal{K}_1 bis \mathcal{K}_3 berechnet werden. Unter Berücksichtigung der durch die Entmischungsstruktur vorgegebenen Abhängigkeit der Ausgangs- von den Eingangssignalen, als

$$\begin{aligned}
 Y_1(\omega) &= -e_2(\varphi)X_1(\omega) + X_2(\omega) \\
 Y_2(\omega) &= e_1(\varphi)X_1(\omega) - X_2(\omega),
 \end{aligned}$$

können die einzelnen Terme ausmultipliziert und so die Statistik der Meßsignale bestimmt werden. Es ergeben sich für die Kostenfunktion fünf zu berechnende Terme, T_1 bis T_5 , mit deren Hilfe die Kreuzkumulante aus Gleichung (A.1) folgende Form annimmt

$$\widehat{\text{cum}}_{22}(\tilde{Y}_1, \tilde{Y}_2) = \frac{1}{T_1 T_2} \cdot (T_5 - T_3 T_3^* - T_4 T_4^*). \quad (\text{B.1})$$

Dabei sind T_1 bis T_5 definiert als

$$T_1 = E[Y_1 Y_1^*] \quad (\text{B.2})$$

$$T_2 = E[Y_2 Y_2^*] \quad (\text{B.3})$$

$$T_3 = E[Y_1 Y_2^*] \quad (\text{B.4})$$

$$T_4 = E[Y_1 Y_2] \quad (\text{B.5})$$

$$T_5 = E[Y_1 Y_1^* Y_2 Y_2^*]. \quad (\text{B.6})$$

Der Vorteil dieser Darstellung ist, daß die einzelnen Terme T_1 bis T_5 unabhängig voneinander ausgewertet werden können.

Berechnung von $T_1 = E[Y_1 Y_1^*] = \sigma_{Y_1}^2$

$$\begin{aligned} E[Y_1 Y_1^*] &= E[(-e_2 X_1 + X_2)(-e_2^* X_1^* + X_2^*)] \\ &= E[|e_2|^2 |X_1|^2 - e_2 X_1 X_2^* - e_2^* X_1^* X_2 + |X_2|^2] \\ &= |e_2|^2 \nu_3 - e_2 \nu_6 - e_2^* \nu_6^* + \nu_4 \\ &= |e_2|^2 \nu_3 - 2\Re(e_2 \nu_6) + \nu_4 \end{aligned} \quad (\text{B.7})$$

Berechnung von $T_2 = E[Y_2 Y_2^*] = \sigma_{Y_2}^2$

$$\begin{aligned} E[Y_2 Y_2^*] &= E[(e_1 X_1 - X_2)(e_1^* X_1^* - X_2^*)] \\ &= E[|e_1|^2 |X_1|^2 - e_1 X_1 X_2^* - e_1^* X_1^* X_2 + |X_2|^2] \\ &= |e_1|^2 \nu_3 - e_1 \nu_6 - e_1^* \nu_6^* + \nu_4 \\ &= |e_1|^2 \nu_3 - 2\Re(e_1 \nu_6) + \nu_4 \end{aligned} \quad (\text{B.8})$$

Berechnung von $T_3 = E[Y_1 Y_2^*]$

$$\begin{aligned} E[Y_1 Y_2^*] &= E[(-e_2 X_1 + X_2) \cdot (e_1^* X_1^* - X_2^*)] \\ &= E[-e_1^* e_2 |X_1|^2 + e_2 X_1 X_2^* + e_1^* X_1^* X_2 - |X_2|^2] \\ &= -e_1^* e_2 \nu_3 + e_2 \nu_6 + e_1^* \nu_6^* - \nu_4 \end{aligned} \quad (\text{B.9})$$

Berechnung von $T_4 = E[Y_1 Y_2]$

$$\begin{aligned} E[Y_1 Y_2] &= E[(-e_2 X_1 + X_2) \cdot (e_1 X_1 - X_2)] \\ &= E[-e_1 e_2 X_1^2 + e_2 X_1 X_2 + e_1 X_1 X_2 - X_2^2] \\ &= -e_1 e_2 \nu_1 + e_2 \nu_5 + e_1 \nu_5 - \nu_2 \end{aligned} \quad (\text{B.10})$$

Berechnung von $T_5 = E[Y_1 Y_1^* Y_2 Y_2^*]$

$$\begin{aligned}
 E[Y_1 Y_1^* Y_2 Y_2^*] &= E[(-e_2 X_1 + X_2)(-e_2^* X_1^* + X_2^*) \cdot (e_1 X_1 - X_2)(e_1^* X_1^* - X_2^*)] \\
 &= E\left[(|e_2|^2 |X_1|^2 - e_2 X_1 X_2^* - e_2^* X_1^* X_2 + |X_2|^2) \right. \\
 &\quad \left. \cdot (|e_1|^2 |X_1|^2 - e_1 X_1 X_2^* - e_1^* X_1^* X_2 + |X_2|^2) \right] \\
 &= E\left[|e_1|^2 |e_2|^2 |X_1|^4 - e_1 |e_2|^2 |X_1|^2 X_1 X_2^* - e_1^* |e_2|^2 |X_1|^2 X_1^* X_2 + |e_2|^2 |X_1|^2 |X_2|^2 \right. \\
 &\quad - |e_1|^2 e_2 X_1 |X_1|^2 X_2^* + e_1 e_2 X_1^2 X_2^{*2} + e_1^* e_2 |X_1|^2 |X_2|^2 - e_2 X_1 X_2^* |X_2|^2 \\
 &\quad - e_2^* |e_1|^2 X_1^* |X_1|^2 X_2 + e_1 e_2^* |X_1|^2 |X_2|^2 + e_1^* e_2^* X_1^{*2} X_2^2 - e_2^* X_1^* X_2 |X_2|^2 \\
 &\quad \left. + |e_1|^2 |X_1|^2 |X_2|^2 - e_1 X_1 X_2^* |X_2| - e_1^* X_1^* X_2 |X_2|^2 + |X_2|^4 \right] \\
 &= |e_1|^2 |e_2|^2 \nu_{11} - e_1 |e_2|^2 \nu_9 - e_1^* |e_2|^2 \nu_9^* + |e_2|^2 \nu_8 - |e_1|^2 e_2 \nu_9 + e_1 e_2 \nu_7 \\
 &\quad + e_1^* e_2 \nu_8 - e_2 \nu_{10} - |e_1|^2 e_2^* \nu_9^* + e_1 e_2^* \nu_8 + e_1^* e_2^* \nu_7^* \\
 &\quad - e_2^* \nu_{10}^* + |e_1|^2 \nu_8 - e_1 \nu_{10} - e_1^* \nu_{10}^* + \nu_{12} \\
 &= 2\Re(e_1 e_2 \nu_7) + 2\Re(e_1 e_2^*) \nu_8 - 2\Re(e_1 |e_2|^2 \nu_9) - 2\Re(|e_1|^2 e_2 \nu_9) - 2\Re(e_2 \nu_{10}) \\
 &\quad - 2\Re(e_1 \nu_{10}) + |e_1|^2 |e_2|^2 \nu_{11} + |e_2|^2 \nu_8 + |e_1|^2 \nu_8 + \nu_{12} \tag{B.11}
 \end{aligned}$$

Die aufwandsreduzierte Kostenfunktion lässt sich mit Hilfe der so berechneten Terme in Abhängigkeit von den Parametern e_1 , e_2 und von den Sensorsignalen X_1 , X_2 formulieren

$$\mathcal{K}(e_1, e_2, X_1, X_2) = \left| \frac{T_5 - |T_3|^2 - |T_4|^2}{T_1 T_2} \right|. \tag{B.12}$$

Betrachtet man $e_1(\omega)$ und $e_2(\omega)$ als Parameter eines zugrundeliegenden Delay-&-Sum-Beamformers, also als reine Phasenfaktoren mit Betrag gleich eins, kann die Kostenfunktion weiter vereinfacht werden.

In ausgeschriebener Form ergibt sich

$$\begin{aligned}
 \mathcal{K}_{\text{Phase}} &= \left| \frac{2\Re(e_1 e_2 \nu_7) + 2\Re(e_1 e_2^*) \nu_8 - 2\Re(e_1 \nu_9) - 2\Re(e_2 \nu_9) - 2\Re(e_2 \nu_{10}) - 2\Re(e_1 \nu_{10})}{(\nu_3 - 2\Re(e_2 \nu_6) + \nu_4) \cdot (\nu_3 - 2\Re(e_1 \nu_6) + \nu_4)} \right. \\
 &\quad \left. + \frac{2\nu_8 + \nu_{11} + \nu_{12} - \left| -e_1^* e_2 \nu_3 + e_2 \nu_6 + e_1^* \nu_6^* - \nu_4 \right|^2 - \left| -e_1 e_2 \nu_1 + e_2 \nu_5 + e_1 \nu_5 - \nu_2 \right|^2}{(\nu_3 - 2\Re(e_2 \nu_6) + \nu_4) \cdot (\nu_3 - 2\Re(e_1 \nu_6) + \nu_4)} \right|.
 \end{aligned}$$

B.2 Aufwandsreduzierter Gradient

Bei der Berechnung des aufwandsreduzierten Gradienten sind zusätzlich zu den in Gleichung (B.2) bis (B.6) definierten Termen noch die Erwartungswerte

$$T_6 = E[Y_1 X_1^*] \quad (\text{B.13})$$

$$T_7 = E[Y_1^* X_1^*] \quad (\text{B.14})$$

$$T_8 = E[Y_2 X_1^*] \quad (\text{B.15})$$

$$T_9 = E[Y_2^* X_1^*] \quad (\text{B.16})$$

$$T_{10} = E[Y_1 Y_1^* Y_2 X_1^*] \quad (\text{B.17})$$

$$T_{11} = E[Y_2 Y_2^* Y_1 X_1^*] \quad (\text{B.18})$$

zu berechnen, mit deren Hilfe sich die partiellen Ableitungen der Kostenfunktion bezüglich der komplexen Parameter e_1^* und e_2^* in folgender Form darstellen lassen:

$$\frac{\partial \widehat{\text{cum}}_{22}(\tilde{Y}_1, \tilde{Y}_2)}{\partial e_1^*} = \frac{1}{T_1 T_2^2} \cdot (T_{10} T_2 - T_5 T_8 - T_6 T_2 T_3^* + |T_3|^2 T_8 - T_4 T_7 T_2 + |T_4|^2 T_8) \quad (\text{B.19})$$

$$\frac{\partial \widehat{\text{cum}}_{22}(\tilde{Y}_1, \tilde{Y}_2)}{\partial e_2^*} = \frac{1}{T_1^2 T_2} \cdot (T_5 T_6 - T_{11} T_1 - |T_3|^2 T_6 + T_3 T_8 T_1 - |T_4|^2 T_6 + T_4 T_9 T_1). \quad (\text{B.20})$$

Die einzelnen Terme T_6 bis T_{11} müssen ausmultipliziert werden, damit der Gradient in Abhängigkeit von der Statistik der Sensorsignale formuliert werden kann.

Auswertung von $T_6 = E[Y_1 X_1^*]$

$$\begin{aligned} E[Y_1 X_1^*] &= E[-e_2 |X_1|^2 + X_1^* X_2] \\ &= -e_2 \nu_3 + \nu_6^* \end{aligned} \quad (\text{B.21})$$

Auswertung von $T_7 = E[Y_1^* X_1^*]$

$$\begin{aligned} E[Y_1^* X_1^*] &= E[-e_2^* X_1^{*2} + X_2^* X_1^* E] \\ &= -e_2^* \nu_1^* + \nu_5^* \end{aligned} \quad (\text{B.22})$$

Auswertung von $T_8 = E[Y_2 X_1^*]$

$$\begin{aligned} E[Y_2 X_1^*] &= E[e_1 |X_1|^2 - X_1^* X_2] \\ &= e_1 \nu_3 - \nu_6^* \end{aligned} \quad (\text{B.23})$$

Auswertung von $T_9 = E[Y_2^* X_1^*]$

$$\begin{aligned} E[Y_2^* X_1^*] &= E[e_1^* X_1^{*2} - X_1^* X_2^*] \\ &= e_1^* \nu_1^* - \nu_5^* \end{aligned} \quad (\text{B.24})$$

Auswertung von $T_{10} = E[Y_1 Y_1^* Y_2 X_1^*]$

$$\begin{aligned} E[Y_1 Y_1^* Y_2 X_1^*] &= E[(-e_2 X_1 + X_2)(-e_2^* X_1^* + X_2^*)(e_1 X_1 - X_2) X_1^*] \\ &= E[e_1 |e_2|^2 |X_1|^4 - |e_2|^2 X_1^* X_2 |X_1|^2 - e_1 e_2 X_1 X_2^* |X_1|^2 + e_2 |X_1|^2 |X_2|^2 \\ &\quad - e_1 e_2^* X_1^* X_2 |X_1|^2 + e_2^* X_1^{*2} X_2^2 + e_1 |X_1|^2 |X_2|^2 - X_1^* X_2 |X_2|^2] \\ &= e_1 |e_2|^2 \nu_{11} - |e_2|^2 \nu_9^* - e_1 e_2 \nu_9 + e_2 \nu_8 - e_1 e_2^* \nu_9^* + e_2^* \nu_7^* + e_1 \nu_8 - \nu_{10}^* \end{aligned} \quad (\text{B.25})$$

Auswertung von $T_{11} = E[Y_1 Y_2 Y_2^* X_1^*]$

$$\begin{aligned} E[Y_2 Y_2^* Y_1 X_1^*] &= E[(e_1 X_1 - X_2)(e_1^* X_1^* - X_2^*)(-e_2 X_1 + X_2) X_1^*] \\ &= E[-|e_1|^2 e_2 |X_1|^4 + |e_1|^2 X_1^* X_2 |X_1|^2 + e_1 e_2 X_1 X_2^* |X_1|^2 - e_1 |X_1|^2 |X_2|^2 \\ &\quad + e_1^* e_2 X_1^* X_2 |X_1|^2 - e_1^* X_1^{*2} X_2^2 - e_2 |X_1|^2 |X_2|^2 + X_1^* X_2 |X_2|^2] \\ &= -|e_1|^2 e_2 \nu_{11} + |e_1|^2 \nu_9^* + e_1 e_2 \nu_9 - e_1 \nu_8 + e_1^* e_2 \nu_9^* - e_1^* \nu_7^* - e_2 \nu_8 + \nu_{10}^* \end{aligned} \quad (\text{B.26})$$

Ebenso wie bei der Berechnung der Kostenfunktion \mathcal{K} in Gleichung (5.41) kann auch der aufwandsreduzierte Gradient bei Annahme reiner Phasenfaktoren vereinfacht werden, da sich die Betragsbildung von e_1 bzw. e_2 erübrigt.

Literaturverzeichnis

- [1] Alijah Ahmed. *Signal Separation*. PhD thesis, University of Cambridge, 2000.
- [2] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems*, volume 8, pages 757–763. The MIT Press, 1996.
- [3] Shun-ichi Amari. Neural Learning in Structured Parameter Spaces – Natural Riemannian Gradient. In Michael C. Mozer, Michael I. Jordan, and Thomas Petsche, editors, *Advances in Neural Information Processing Systems*, volume 9, page 127. The MIT Press, 1997.
- [4] Jörn Anemüller. *Across-Frequency Processing in Convolutional Blind Source Separation*. PhD thesis, Universität Oldenburg, 2001.
- [5] S. P. Applebaum and D. J. Chapman. Adaptive Arrays with main beam constraints. *IEEE Trans. on AP*, AP-24:650–662, September 1976.
- [6] H. B. Barlow, T. P. Kaushal, and G. J. Mitchison. Finding Minimum Entropy Codes. *Neural Computation*, 1:412–423, 1989.
- [7] W. Baumann, D. Kolossa, and R. Orglmeister. Maximum likelihood permutation correction for convolutional source separation. In *ICA 2003*, pages 373–378, Nara, Japan, Apr 2003.
- [8] Wolf Baumann, Bert-Uwe Köhler, Dorothea Kolossa, and Reinhold Orglmeister. Real time separation of convolutional mixtures. In *ICA 2001*, pages 65–69, 2001.
- [9] A. J. Bell and T. J. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1004–1034, 1995.

- [10] Adel Belouchrani, Karim Abed-Meraim, Jean-Francois Cardoso, and Eric Moulines. A Blind Source Separation Technique Using Second-Order Statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444, Februaray 1997.
- [11] Stefan Harmeling (GMD First Berlin). http://ida.first.gmd.de/harmeli/download/download_convbss.html, 2001.
- [12] E. Bingham and A. Hyvärinen. A Fast Fixed-Point Algorithm for Independent Component Analysis of Complex Valued Signals. *International Journal of Neural Systems*, 10(1):1–8, 2000.
- [13] J. Bitzer, K. U. Simmer, and K. D. Kammeyer. Multi-microphone noise reduction by post-filter and superdirective beamformer. In *Proc. Int. Workshop Acoust. Echo and Noise Control*, pages 100–103, Pocono Manor, USA, September 1999.
- [14] Joerg Bitzer and K. Uwe Simmer. *Superdirective Microphone Arrays*, chapter 2, pages 19–38. In Brandstein and Ward [17], 2001.
- [15] Tobias Blaschke and Laurenz Wiskott. Cubica: Independent component analysis by simultaneous third- and fourth-order cumulant diagonalization. *Computer Science Preprint Server (CSPS) Computational Intelligence*, April 2003.
- [16] R. Le Bouquin and G. Faucon. Using the coherence function for noise reduction. *IEE-Proceedings I*, 139(3):276–280, June 1992.
- [17] M. Brandstein and D. Ward, editors. *Microphone Arrays*. Springer, Berlin Heidelberg New York, 2001.
- [18] D.H. Brandwood. A complex gradient operator and its application in adaptive array theory. *IEE-Proceedings F and H*, 130(1):11–16, Feb 1983.
- [19] H. Buchner, R. Aichner, and W. Kellermann. A Generalization of a Class of Blind Source Separation Algorithms for Convolutive Mixtures. In *ICA03*, pages 945–950, Nara, Japan, Apr 2003.
- [20] V. Capdevielle, Ch. Serviere, and J.L. Lacoume. Blind separation of wide-band sources in the frequency domain. In *Proceedings ICASSP 1995*, pages 2080– 2083, 1995.
- [21] J.-F. Cardoso. Fourth-order cumulant structure forcing, application to blind array processing. In *Proc. IEEE SP Workshop on SSAP*, pages 136–139. IEEE, October 1992.

- [22] J.-F. Cardoso. Estimating equations for source separation. In *Proc. ICASSP'97*, pages 3449–3452, Munich, 1997.
- [23] J.-F. Cardoso. High order contrasts for independent component analysis. *Neural Computation*, 11:157–192, 1999.
- [24] J.-F. Cardoso and Antoine Soudoumiac. Blind Beamforming for Non Gaussian Signals. *IEE Proceedings F*, 140(6):362–370, December 1993.
- [25] Jean-Francois Cardoso. Blind Signal Separation: Statistical Principles. *Proc. of the IEEE*, 86(10):2009–2025, October 1998.
- [26] P. Comon. Independent component analysis, a new concept? *Signal Processing, Elsevier*, 36(3):287–314, 1994.
- [27] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.
- [28] I. Mc Cowan and H. Bourslard. Microphone Array Post-Filter for Diffuse Noise Field. In *Proc. ICASSP 2002*, volume 1, pages 905–908, Florida, 2002.
- [29] Henry Cox, Robert M. Zeskind, and Mark M. Owen. Robust Adaptive Beamforming. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 35(10):236–243, October 1987.
- [30] S. Ding, T. Hikichi, T. Niitsuma, M. Hamatsu, and K. Sugai. Recursive Method for Blind Source Separation and its Applications to Real-Time Separations of Acoustic Signals. In *ICA03*, pages 517–522, Nara, Japan, Apr 2003.
- [31] Matthias Dörbecker. *Mehrkanalige Signalverarbeitung zur Verbesserung akustisch gestörter Sprachsignale am Beispiel elektronischer Hörhilfen*. PhD dissertation, RWTH Aachen, 1998.
- [32] Martin Drews. *Mikrofonarrays und mehrkanalige Signalverarbeitung zur Verbesserung gestörter Sprache*. PhD dissertation, Technische Universität Berlin, 2000.
- [33] Eberhard Hänsler. *Statistische Signale: Grundlagen und Anwendungen*. Springer-Verlag, Berlin, Heidelberg, New York, 2 edition, 1997.
- [34] F. Ehlers and H. G. Schuster. Blind Separation of Convolutional Mixtures and an Application in Automatic Speech Recognition in a Noisy Environment. *IEEE Transaction on Signal Processing*, 45(10):2608–2612, 1997.

- [35] Gary W. Elko. *Spatial Coherence Functions for Differential Microphones in Isotropic Noise fields*, chapter 4, pages 62–85. In Brandstein and Ward [17], 2001.
- [36] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Trans. Acoust., Speech, Signal Processing*, 32(6):1109–1121, 1984.
- [37] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Trans. Acoust., Speech, Signal Processing*, 33(2):443–445, 1985.
- [38] Otis Lamont Frost. An Algorithm for Linearly Constrained Adaptive Array Processing. *Proceedings of the IEEE*, 60(8):926–935, 1972.
- [39] Björn Geipel. Realzeitfähige Blinde Quellentrennung gefalteter Mischungen mittels digitalem Signalprozessor. Diplomarbeit, Technische Universität Berlin, 2004.
- [40] E. N. Gilbert and S. P. Morgan. Optimum Design of Directive Antenna Arrays Subject to Random Variations. *The Bell System Technical Journal*, 34:637–663, 1955.
- [41] Daniel W. Griffin and Jae S. Lim. Signal Estimation from Modified Short-Time Fourier Transform. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 32(2):236–243, April 1984.
- [42] L. J. Griffiths and C. W. Jim. An Alternative Approach to Linearly Constrained Adaptive Beamforming. *IEEE Trans. on Antennas and Propagation*, 30(1):27–34, January 1982.
- [43] Simon Haykin, editor. *Unsupervised Adaptive Filtering Volume 1: Blind Source Separation*. John Wiley, New York, 2000.
- [44] Simon Haykin. *Adaptive Filter Theory*. Prentice-Hall, Upper Saddle River, New Jersey, 4th edition, 2002.
- [45] <http://morph ldc.upenn.edu/Catalog/LDC93S10.html>. Linguistik Data Consortium - TIDIGITS.
- [46] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

- [47] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley, New York, 2001.
- [48] A. Hyvärinen and E. Oja. A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation*, 9:1483–1492, 1997.
- [49] Don H. Johnson and Dan E. Dudgeon. *Array Signal Processing: Concepts and Techniques*. Prentice Hall, Englewood Cliffs, 1993.
- [50] Bert-Uwe Köhler. *Konzepte der statistischen Signalverarbeitung*. Springer, Berlin, 2005.
- [51] Mirko Knaak. *Rekonstruktion gestörter Maschinengeräusche durch mehrkanalige Signalverarbeitung*. PhD thesis, Technische Universität Berlin, 2003.
- [52] Mirko Knaak and Dieter Filbert. Acoustical semi-blind source separation for machine monitoring. In *ICA 2001*, pages 361–366, 2001.
- [53] Bert-Uwe Köhler. *Realzeitfähige blinde Quellentrennung am Beispiel elektroenzephalographischer Signale*. PhD thesis, Technische Universität Berlin, 1999.
- [54] D. Kolossa and R. Orglmeister. Nonlinear postprocessing for blind speech separation. In *ICA 2004*, pages 832–839, 2004.
- [55] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura. Blind signal separation using directivity pattern. Technical report, Japanese Society for Artificial Intelligence, November 1999.
- [56] Satoshi Kurita, Hiroshi Saruwatari, Shoji Kajita, and Fumitada Itakura. Evaluation of Blind Signal Separation Method using Directivity Pattern Under Reverberant Conditions. *IEEE Trans. on Speech and Audio Processing*, pages 320–327, May 2000.
- [57] Heinrich Kuttruff. *Room acoustics*. Spon Press, London, 4th edition, 2000.
- [58] Russel H. Lambert. *Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath Mixtures*. PhD thesis, University of Southern California, May 1996.
- [59] S. Laughlin. A simple coding procedure enhances a neuron’s information capacity. *Zeitschrift für Naturforschung*, 36(C):910–912, 1981.

- [60] T.-W. Lee, A. Ziehe, R. Orglmeister, and T.J. Sejnowski. Combining Time Delayed Decorrelation and ICA: Towards Solving the Cocktail Party Problem. In *Proc. ICASSP*, volume 2, pages 1249–1252, 1998.
- [61] Te-Won Lee, Anthony Bell, and Reinhold Orglmeister. Blind source separation of real world signals. In *Int. Conf. on Neural Networks*, pages 564–569, Houston, 1997.
- [62] R. Linsker. Self-Organization in a Perceptual Network. *Computer*, 21:105–117, 1988.
- [63] R. Linsker. An application of the principle of maximum information preservation to linear systems. *Advances in Neural Information Processing Systems*, 1, 1989.
- [64] Vijay K. Madisetti and Douglas B. Williams, editors. *The Digital Signal Processing Handbook*. CRC Press LLC, Boca Raton, Florida, 1998.
- [65] A. Mansour and C. Jutten. Fourth-order criteria for blind sources separation. *IEEE Transactions on Signal Processing*, 43(8):2022 – 2025, August 1995.
- [66] K. Matsuoka, M. Ohya, and M. Kawamoto. A neural net for blind separation of nonstationary signals. *Neural Networks*, 8(3):411–419, 1995.
- [67] Cleve Moler. *Numerical Computing with Matlab*. THE MATHWORKS INC. <http://www.mathworks.com/moler/>, 2004.
- [68] L. Molgedey and H. G. Schuster. Separation of a Mixture of Independent Signals Using Time Delayed Correlations. *Physical Review Letters*, 72:3634–3637, 1994.
- [69] R. Monzingo and T. Miller. *Introduction to Adaptive Arrays*. Wiley and Sons, New York, 1980.
- [70] N. Murata, S. Ikeda, and A. Ziehe. An approach to blind source separation based on temporal structure of speech signals. *BSIS Tech. Report*, 1998.
- [71] Chrysostomos L. Nikias and Athina P. Petropulu. *Higher-Order Spectra Analysis*. PTR Prentice Hall, Englewood Cliffs, New Jersey, 1993.
- [72] Markos Papageorgiou. *Optimierung*. Oldenbourg, 1991.
- [73] A. Papoulis. *Probability, random variables and stochastic processes*. McGraw-Hill, New York, 2nd edition, 1984.
- [74] L. Parra and C. Spence. Convolutional blind source separation of non-stationary sources. *IEEE Trans. on Speech and Audio Processing*, pages 320–327, May 2000.

- [75] Lucas Parra and Christopher Alvino. Geometric source separation: Merging convolutive source separation with geometric beamforming. *IEEE Transactions on Speech and Audio Processing*, 10(6):352–362, September 2002.
- [76] Jean-Christophe Pesquet and Eric Moreau. Cumulant-based independence measures for linear mixtures. *IEEE Transactions on Information Theory*, 47(5):1947 – 1956, July 2001.
- [77] S. Unnikrishna Pillai. *Array signal processing*. Springer, Berlin Heidelberg New York, 1989.
- [78] Scott Rickard, Radu Balan, and Justinian Rosca. Real-time time-frequency based blind source separation. In *ICA 2001*, pages 651–656, 2001.
- [79] P. Rombach, M. Mullenborn, U. Klein, and K. Rasmussen. The first low voltage, low noise differential silicon microphone, technology development and measurement results. In *The 14th IEEE International Conference on Micro Electro Mechanical Systems, MEMS2001*, pages 42–45, January 2001.
- [80] Hiroshi Saruwatari, Toshiya Kawamura, and Kiyohiro Shikano. Blind Source Separation Based on Fast-Convergence Algorithm Using ICA and Array Signal Processing. In *ICA 2001*, pages 412–417, 2001.
- [81] D. Schobben, K. Torkkola, and P. Smaragdis. Evaluation of blind signal separation methods. In *ICA 1999*, pages 261–266, Aussois, Jan. 1999.
- [82] Daniel W.E. Schobben. *Real-Time Adaptive Concepts in Acoustics*. Kluwer Academic Publishers, Dordrecht, 2001.
- [83] Hans Rudolf Schwarz. *Numerische Mathematik*. B. G. Teubner Stuttgart, 1997.
- [84] Dae C. Shin and Chrysostomos L. Nikias. Complex ambiguity function based on fourth-order statistics for joint estimation of frequency-delay and time-delay of arrival. In *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 461–465, November 1-3 1993.
- [85] K. Uwe Simmer, Joerg Bitzer, and Claude Marro. *Post-Filtering Techniques*, chapter 3, pages 39–60. In Brandstein and Ward [17], 2001.

- [86] I. Sliméne, S. Ben Jebara, and A. Benazza-Benyahia. Use of coherence in wavelet and multiwavelet domain for microphone array speech enhancement. In *First International Symposium on Control, Communications and Signal Processing*, pages 627–630, March 21-24 2004.
- [87] M.M. Sondhi. An adaptive echo canceller. *Bell Syst. Tech. J.*, 46:497–511, 1966.
- [88] Carsten Sydow. Broadband beamforming for a microphone array. *Journal of the Acoustical Society of America*, 96(2):845–849, August 1994.
- [89] *Tafelwerk Mathematik, Physik, Chemie: Klassen 11/12*. Volk und Wissen Volkseigener Verlag Berlin, Berlin, 1982.
- [90] W. Täger. Near field superdirectivity (NFSD). In *Proceedings of ICASSP '98*, pages 2045–2048, 1998.
- [91] Charles W. Therrien. *Discrete Random Signals and Statistical Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1992.
- [92] Quifeng Thu. The Effect of Additive Noise on Speech Amplitude Spectra: A Quantitative Analysis. *IEEE Signal Processing Letters*, 9(9):275–277, September 2002.
- [93] Harry L. Van Trees. *Optimum Array Processing*. John Wiley & Sons, New York, 2002.
- [94] Jitendra K. Tugnait. On time delay estimation with unknown spatially correlated gaussian noise using fourth-order cumulants and cross cumulants. *IEEE Transactions on Signal Processing*, 39(6), June 1991.
- [95] P. Vary, U. Heute, and W. Hess. *Digitale Sprachsignalverarbeitung*. B.G. Teubner Stuttgart, 1998.
- [96] Saeed V. Vaseghi. *Advanced Signal Processing and Digital Noise Reduction*. John Wiley and B.G. Teubner, 1996.
- [97] Barry D. Van Veen and Kevin M. Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Magazine*, 5(2):4–24, April 1988.
- [98] Barry Van Veen and Kevin M. Buckley. *Beamforming Techniques for Spatial Filtering*, chapter 61, pages 1–21. In Madisetti and Williams [64], 1998.

- [99] E. Visser and T.-W. Lee. Application of Blind Source Separation in Speech Processing for Combined Interference Removal and Robust Speaker Detection using a Two-Microphone Setup. In *ICA03*, pages 325–329, Nara, Japan, Apr 2003.
- [100] Khurram Waheed and Fathi M. Salam. A data-derived quadratic independence measure for adaptive blind source recovery in practical applications. In *45th IEEE International Midwest Symposium on Circuits and Systems, Tulsa, Oklahoma*, volume 3, pages 473–476, August 4-7 2002.
- [101] Darren B. Ward, Rodney A. Kennedy, and Robert C. Williamson. Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns. *Journal of the Acoustical Society of America*, 97(2):1023–1034, Feb 1997.
- [102] Darren B. Ward, Rodney A. Kennedy, and Robert C. Williamson. *Constant Directivity Beamforming*, chapter 1, pages 3–17. In Brandstein and Ward [17], 2001.
- [103] <http://www.dspalgorithms.com/download.html>, 2004.
- [104] Webseite DaimlerChrysler AG. <http://www.mercedes-benz.de>, 2004.
- [105] B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode. Adaptive antenna systems. *Proceedings IEEE*, 55:2143–2159, December 1967.
- [106] Bernard Widrow and Samuel D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, Upper Saddle River, New Jersey, 1985.
- [107] R. Zelinski. Ein Geräuschunterdrückungssystem mit zweidimensionaler Mikrophonengruppe und nachgeschalteter adaptiver Wiener Filterung. In *6. Aachener Symposium für Signaltheorie*, pages 372–375. Springer Verlag, 1987.
- [108] Rainer Zelinski. Mikrofon-Arrays mit superdirektiven Eigenschaften zur Sprachsignalübertragung. *Frequenz*, 50(9-10):198–204, 1996.