

Rekonstruktion gestörter Maschinengeräusche durch mehrkanalige Signalverarbeitung

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

vorgelegt von
Diplom Ingenieur
Mirko Knaak
aus Berlin

Berlin, 2003
D83

Promotionsausschuß:

Vorsitzender: Prof. Dr. rer. nat. Klaus Obermayer

1. Gutachter: Prof. Dr.-Ing. Dieter Filbert

2. Gutachter: Prof. Dr. rer. nat. Andreas Wendemuth
(Otto-von-Guericke-Universität Magdeburg)

Tag der wissenschaftlichen Aussprache: 10.10.2003

Danksagung

Einen besonderen Dank für die wegweisenden Ideen und den fruchtbaren Gedankenaustausch möchte ich Prof. Dr.-Ing. Dieter Filbert aussprechen. Insbesondere förderte er meinen Eintritt in die wissenschaftlichen Community dieses Forschungsgebietes. Prof. Dr. rer. nat. Andreas Wendemuth hat freundlicherweise die Mühen des Zweitgutachtens übernommen. Dafür danke ich ihm sehr.

An dieser Stelle möchte ich ganz besonders dem Speech Open Lab der NTT Corp. und seinem Senior Researcher Dr. Shoji Makino für die Einladung danken, in der inspirierenden Atmosphäre dieses Labs meine Forschungen fortzusetzen. Ohne sie und die unendlich spannenden Diskussionen wäre der CICA-Algorithmus sicher nie entstanden: Araki-san, Makino-san, Sawada-san, Mukai-san - Arigatou!

Bedanken möchte ich mich bei allen Mitarbeitern und Studenten des Fachgebiets Elektrische Meßtechnik und befreundeter Fachgebiete, die zum Gelingen dieser Arbeit in vielerlei Hinsicht beigetragen haben: Jan-Mark Batke, Uzme Bazarsuren, Sebastian Knorr, Dorothea Kolossa, Matthias Kunter, Stefan Lachmann, Michael Lindemann, Tian Ying Li, Domingo Mery und Manfred Winkelkemper.

Weiterhin danke ich Michael Fausten und Joseph Mainka, die mir die Möglichkeit gegeben haben, Messungen direkt an einer Produktionslinie aufnehmen zu dürfen.

Ferner bedanke ich mich bei allen Mitgliedern des Fachgebiets, insbesondere bei Lisa Kraft für das angenehme Betriebsklima und Wolfgang Mahnfitz für seine Hilfsbereitschaft bei allen Problemen (nicht nur) rund um den Rechner.

Meine Eltern haben mich erst zu dem gemacht, was ich bin. Sie unterstützten mich durch mein gesamtes Studium in allem. Dafür danke ich ihnen von ganzem Herzen.

Für die große Hilfe beim Redigieren der Arbeit und das ständige „Für-einander-da-sein“ während der gesamten Forschungszeit danke ich meiner lieben Freundin Astrid Freier. Damit half sie mir nicht nur bei der anstrengenden Fertigstellung sondern auch in allen Tiefpunkten, in denen die Algorithmen einfach nicht konvergieren wollten. Nicht nur dadurch hat sie mir mehr gegeben, als sich hier ausdrücken läßt.

Berlin, zu Weihnachten 2003

Mirko Knaak

Zusammenfassung

Mit der vorliegenden Arbeit wird das Ziel verfolgt, die Independent Component Analysis (ICA) als Vorverarbeitung für die akustische Güteprüfung nutzbar zu machen. Die Grundidee der ICA besteht darin, mehrkanalige Signale in ihre statistisch unabhängigen Komponenten zu zerlegen. Da die Trennung blind, d.h. ohne Vorwissen über die Mischung oder die Quellen erfolgt, wird sie bei Mischungen von Luftschall als akustische blinde Quellentrennung bezeichnet.

Bei der Anwendung der akustischen blinden Quellentrennung bestehen zur Zeit noch wesentliche Hürden. Zu ihnen gehört die Limitierung der aktuell verfügbaren Algorithmen auf zwei Quellen, die getrennt werden können. Damit darf das Nutzsinal nur von einem Störsignal überlagert sein, was für die praktische Anwendung einschränkend ist, weil die Anzahl der relevanten Störquellen oft größer als zwei ist und man sie a priori nicht bestimmen kann.

Um neue Algorithmen zu entwickeln und die bisherigen Limitierungen besser zu verstehen, wird die akustische Blinde Quellentrennung dazu im Kontext der mehrkanaligen akustischen Signalverarbeitung theoretisch analysiert. Das bemerkenswerte Ergebnis dieser Analyse ist, daß die klassischen adaptiven Beamformer und die akustische blinde Quellentrennung grundsätzlich äquivalent sind: Die N-kanalige akustische blinde Quellentrennung besteht aus N adaptiven Minimum-Variance-Beamformern, wobei die Einfallsrichtung (DOA) des Nutzsinalns implizit aus den Signalen geschätzt wird. Aus dieser Erkenntnis wird ein neuer Algorithmus hergeleitet, der die ICA nicht mehr blind, sondern mit Nebeninformationen anwendet. Für ihn werden sowohl Vorteile gegenüber der blinden Quellentrennung als auch eine deutlich geringere Parameterempfindlichkeit gegenüber den adaptiven Beamformern nachgewiesen.

Weiterhin wird ein mehrkanaliger Algorithmus vorgeschlagen, der auf einem Subspace-Ansatz basiert. Er findet Zeitintervalle relativer Ruhe, in denen dann das Nutzsinal zuverlässig klassifiziert werden kann. Ein derartiges Vorgehen ist zwar nur nutzbar, wenn die Störquellen intermittierend vorkommen. Dafür bietet der Algorithmus in diesem Spezialfall eine sehr schnelle, kostengünstige Lösung des Problems. Schließlich wird eine Lösung eines der wichtigsten Probleme der ICA, das Permutationsproblem, vorgeschlagen, die eine charakteristische Eigenschaft von Maschinensignalen, nämlich die annähernde Periodizität, ausnutzt.

Experimentelle Tests in definierten Testumgebungen zeigen abschließend die Anwendbarkeit der neuen Algorithmen für die Verbesserung der Klassifikationsergebnisse.

Inhaltsverzeichnis

Danksagung	iii
Zusammenfassung	v
Inhaltsverzeichnis	vi
1 Einleitung und Zielsetzung	1
2 Akustische Qualitätskontrolle	5
3 Mehrkanalige Schallmessung	11
3.1 Ausbreitung der Schallsignale im Raum	11
3.1.1 Entstehung und Ausbreitung von Schallwellen	11
3.1.2 Wellenbeschreibung im Frequenz-Wellenzahlbereich	14
3.2 Einführung in Mikrofon-Arrays	16
3.2.1 Räumlich ausgedehnte Signalaufnahme	16
3.2.2 Räumliche diskrete Signalaufnahme	19
4 Beamforming	21
4.1 Definitionen und Annahmen über die Signale	22
4.2 Konventionelle Beamformer	26
4.2.1 Delay-&-Sum-Beamformer	27
4.2.2 Filter-&-Sum-Beamformer	29
4.3 Adaptive Beamformer	31
4.3.1 Minimum Variance Beamformer	32
4.3.2 Maximum Likelihood Beamformer	35
4.4 Einfluß einer ungenauen Schätzung des Verzögerungsvektors	37
4.4.1 Gründe für die ungenaue Schätzung	37
4.4.2 Konsequenzen für den ML-Beamformer	40

4.4.3	Konsequenzen für den MV-Beamformer	42
4.4.4	Möglichkeiten für einen robusten Beamformer	47
5	Blinde Trennung akustischer Mischungen	49
5.1	Idee der blinden Quellentrennung	50
5.1.1	Darstellung der statistischen Unabhängigkeit	50
5.1.2	Mischungs- und Entmischungsmodell für akustische Mischungen	52
5.1.3	ICA im Frequenzbereich	54
5.1.4	Hauptachsentransformation	56
5.2	Nichtlineare Dekorrelation	58
5.2.1	Maximierung der Negentropie (FastICA)	60
5.2.2	Maximierung der Transinformation (Infomax)	64
5.2.3	Minimierung der Kreuzkumulanten (Jade)	70
5.3	Kontrastfunktionen auf Basis der Statistik zweiter Ordnung (SOBI) .	72
5.4	Behandlung des Permutations- und Skalierungsproblems	77
5.4.1	Allgemeine Methoden zur Behandlung des Permutationsproblems	77
5.4.2	Ausnutzung der Eigenschaften der Motorsignale	80
5.4.3	Skalierungsproblem	84
6	Vergleich adaptive Beamformer und blinde Quellentrennung	87
6.1	Äquivalenz zwischen adaptiven Beamformern und konvolutiver BSS .	88
6.2	ICA mit geometrischen Randbedingungen	90
6.2.1	Ableitung des Algorithmus	90
6.2.2	Simulationen	93
6.2.3	Analytische Untersuchung der Konvergenz	94
7	Anwendung auf Maschinensignale	99
7.1	Voraussetzung für die stochastischen Algorithmen	100
7.2	Ausnutzung der Diskontinuität der Störsignale	104
7.2.1	Schätzung des Signalunterraumes	105
7.2.2	Bestimmung der Dimensionalität des Signalunterraums	107
7.2.3	Untersuchungsbedingungen und Algorithmus	109

8	Untersuchung der neuen Verfahren und Ergebnisse	111
8.1	Beschreibung der Testumgebungen	113
8.2	Bewertungsmaße für die Rekonstruktion	114
8.2.1	Maße für die Unterdrückung der Nebengeräusche (SIR)	115
8.2.2	Maße für die Qualität des rekonstruierten Signals (Klassifizierbarkeit)	117
8.3	Ergebnisse der ICA-Verfahren	122
8.3.1	Auswahl der Verfahren	122
8.3.2	Quantitativer Vergleich	124
8.4	Ergebnisse des Subspace-Algorithmus	128
9	Zusammenfassung	135
	Abbildungsverzeichnis	137
	Tabellenverzeichnis	141
	Literaturverzeichnis	142

Kapitel 1

Einleitung und Zielsetzung

Essentieller Bestandteil einer qualitätsgerechten Produktion ist die Endkontrolle möglichst aller produzierten Güter. Bei elektrischen Geräten ist neben dem Funktionstest und der Überprüfung der elektrischen Eigenschaften auch eine akustische Güteprüfung zentraler Teil der Endkontrolle, denn die Geräuscharmheit gehört zu den wichtigsten Kundenwünschen für solche Produkte. Darüber hinaus lassen sich am Klang zahlreiche Fertigungs- und Montagefehler erkennen, die sich in anderen Größen nicht abzeichnen.

Derzeit erfolgt die akustische Güteprüfung fast ausschließlich manuell. Zu den Gründen gehören die erheblichen Nebengeräusche in einer Fertigungshalle. Sie sind sehr vielfältiger Natur, übersteigen oft den Pegel der zu untersuchenden Maschinen, machen die automatische Prüfung unzuverlässig und verursachen hohe Kosten für die Schalldämmung der Prüfstände. *Ihre Unterdrückung mit Methoden der Signalverarbeitung ist der Gegenstand dieser Arbeit.*

Die theoretische Basis für die Unterdrückung ist die Independent Component Analysis (ICA). Ihre Grundidee besteht darin, mehrkanalige Signale in ihre statistisch unabhängigen Komponenten zu zerlegen. Die ICA wurde Anfang der 1990er Jahre u.a. durch Christian Jutten, Jeanny Hérault, Jean-François Cardoso, Erkki Oja, Shun-ichi Amari, Antony Bell und Terence Sejnowski aus Erkenntnissen der neuronalen Informationsverarbeitung und der Statistik höherer Ordnung entwickelt. Ihre Hauptanwendung ist die Blinde Quellentrennung (BSS): Die Rekonstruktion von Signalen aus ihrer linearen Mischung¹ ohne die Nutzung weiterer Informationen. Seit ihrer Einführung werden die wissenschaftlichen Bereiche, in denen die BSS eingesetzt wird, immer zahlreicher. Sie reichen von der Medizintechnik, über die Telekommunikation, die Sprachverarbeitung, die Geologie, die Seismologie und die Bildverarbeitung bis hin zu volkswirtschaftlichen und finanzmathematischen Bereichen.

¹Unter linearer Mischung wird in dieser Arbeit die additive Überlagerung zweier oder mehrerer Signale mit unterschiedlichen Verstärkungsfaktoren oder Zeitverzögerungen verstanden. Es handelt sich dabei *nicht* wie in der Nachrichtenübertragung um eine multiplikative Überlagerung.

Seitdem es Te-Won Lee 1997 gelungen ist, mit der Blinden Quellentrennung eine Lösung für den bekannten Cocktail-Party-Effekt² - wenn auch in einem *sehr* einfachen Fall - zu präsentieren, gehört auch die Akustik zu den Anwendungsgebieten. Die Rekonstruktion der Quellensignale ist in der Akustik jedoch besonders schwierig, weil im Gegensatz zu den meisten anderen Bereichen eine sog. konvolutive Mischung vorliegt: Die zu trennenden Signale sind mit der Raumimpulsantwort gefiltert und dann aufaddiert.

Da die akustische Blinde Quellentrennung ein mehrkanaliges Verfahren ist, müssen die gemischten Signale mit mehreren Kanälen aufgenommen werden, d.h. es ist ein Mikrofon-Array nötig. Bei ihrer Anwendung in der akustischen Güteprüfung bestehen zur Zeit noch wesentliche Hürden. Zu ihnen gehört die Limitierung der aktuell verfügbaren Algorithmen auf zwei Quellen, die getrennt werden können. Damit darf das Nutzsignal nur von einem Störsignal überlagert sein, was für die praktische Anwendung ausgesprochen einschränkend ist, weil die Anzahl der relevanten Störquellen oft größer als zwei ist und man sie a priori nicht bestimmen kann.

Mit der vorliegenden Arbeit wird das Ziel verfolgt, die ICA als Vorverarbeitung für die akustische Güteprüfung nutzbar zu machen. Um neue Algorithmen zu entwickeln und die bisherigen Limitierungen besser zu verstehen, wird die akustische Blinde Quellentrennung dazu im Kontext der mehrkanaligen akustischen Signalverarbeitung theoretisch analysiert, was bis vor kurzem durch das andere Herkunftsgebiet der ICA (Statistik) nur ungenügend getan wurde. Basierend auf den hieraus resultierenden Erkenntnissen wurden neue Algorithmen entwickelt, die die ICA nicht mehr blind, d.h. ohne Vorwissen, einsetzen und gleichzeitig die Einsetzbarkeit verbessern.

Die Arbeit hat folgende Gliederung: Nach dieser Einleitung folgt in Kap. 2 ein Überblick über die Klassifikation elektrischer Kleinmotoren und Haushaltsgeräte. In Kap. 3 werden die wichtigsten Grundlagen der Signalaufnahme mit einem Mikrofon-Array, d.h. mit einem diskreten, ausgedehnten Sensor, gelegt. Die Untersuchung der in der Akustik üblichen mehrkanaligen Verfahren, der sog. Beamformer, erfolgt in Kap. 4. Dabei wird herausgearbeitet, daß die adaptiven Beamformer durch ihre hohe Parameterempfindlichkeit nicht für die akustische Güteprüfung geeignet sind.

Kap. 5 befaßt sich mit der ICA und ihrer Anwendung auf die Trennung konvolutiver Mischungen und vermittelt einen Überblick über die aktuellen Verfahren. Für eines der wichtigsten Probleme, das Permutationsproblem, wird eine neue Lösung vorgeschlagen, die eine charakteristische Eigenschaft von Maschinensignalen, nämlich die annähernde Periodizität, ausnutzt.

Der Schwerpunkt von Kap. 6 ist die theoretische Betrachtung über die Wirkungsweise der ICA in der Akustik. Das bemerkenswerte Ergebnis ist, daß die adaptiven Beamformer und die konvolutive blinde Quellentrennung grundsätzlich äquivalent sind: Die N-kanalige konvolutive blinde Quellentrennung besteht aus N adaptiven

²Der Cocktail-Party-Effekt wurde von Cherry und Collin (1953) erstmals formuliert und beschreibt die bemerkenswerte menschliche Fähigkeit, etwa bei einer Cocktailparty oder anderen Anlässe verschiedene Sprachsignale zu trennen. Das menschliche Gehirn/Ohr wendet vermutlich eine Kombination zahlreicher Methoden an und verbindet die Trennung mit der Erkennung des Sprachinhalts.

Minimum-Variance-Beamformern, wobei die Einfallsrichtung (DOA) des Nutzsignals implizit aus den Signalen geschätzt wird. Mit dieser Erkenntnis wird ein neuer Algorithmus vorgeschlagen, der die ICA nicht mehr blind, sondern mit Nebeninformationen anwendet. Für ihn werden sowohl Vorteile gegenüber der blinden Quellentrennung als auch eine deutlich geringere Parameterempfindlichkeit gegenüber den adaptiven Beamformern nachgewiesen.

Mit der Optimierung der zuvor besprochenen Algorithmen beschäftigt sich Kap. 7. Zusätzlich wird in diesem Kapitel ein weiterer mehrkanaliger Algorithmus vorgeschlagen, der auf einem Subspace-Ansatz basiert. Er findet Zeitintervalle relativer Ruhe, in denen dann das Nutzsignal zuverlässig klassifiziert werden kann. Ein derartiges Vorgehen ist zwar nur nutzbar, wenn die Störquellen intermittierend vorkommen. Dafür bietet dieser Algorithmus aber eine sehr schnelle, kostengünstige Lösung des Problems selbst bei starkem Diffusschall an.

Experimentelle Tests in definierten Testumgebungen folgen in Kap. 8. Abschließend faßt Kap. 9. die Arbeit unter Berücksichtigung der vorgestellten neuen Erkenntnisse und Algorithmen zusammen.

Kapitel 2

Akustische Qualitätskontrolle

Die akustische Qualitätskontrolle beschäftigt sich damit, aus dem Geräusch einer Maschine auf ihren Qualitätszustand zu schließen. Im folgenden wird nun genauer erläutert und dargestellt, wie die in dieser Arbeit vorgestellten Verfahren in die akustische Güteprüfung integriert werden.

Untersuchungsobjekte

Die Untersuchungsobjekte der vorliegenden Arbeit sind elektrische Kleinmotoren und Haushaltsgeräte, insbesondere Waschmaschinen.

Elektrische Kleinmotoren sind Kommutator-Reihenschlußmaschinen sowie Gleichstrommaschinen mit Permanenterregung¹ mit einer Leistung bis 1 kW. Sie werden häufig in Kraftfahrzeugen, Haushaltsgeräten und industriellen Steuerungen eingesetzt. In Kraftfahrzeugen sind es permanent erregte Gleichstrommaschinen und in Haushaltsgeräten etc. Universalmotoren. Das Untersuchungsgebiet ergibt sich zum einen daraus, daß diese Motoren auch derzeit akustisch (wenn auch manuell) auf Fehler überprüft werden. Zum anderen ist ihre effiziente Qualitätskontrolle von großer wirtschaftlicher Bedeutung: Kleinmotoren machen mit ca. 5 Mrd. EUR etwa 38% des Umsatzes aller elektrischen Antriebe in Deutschland aus, und ihre Produktion ist trotz Konjunkturschwäche von 1995 bis 2002 stetig gestiegen, während die der größeren Motoren stagnierte.

Obige Ausführungen gelten auch für elektrische Haushaltsgeräte. Sie machen 45% des Umsatzes von elektrotechnischen Gebrauchsgütern aus und generieren einen Umsatz von ca. 10 Mrd. EUR, was die wirtschaftliche Bedeutung dieses Segments zeigt. Waschmaschinen, die hier schwerpunktmäßig untersucht werden, sind dabei der größte Umsatztreiber².

Die Untersuchungsobjekte werden einer hundertprozentigen Prüfung unterzogen. Obwohl dies bei den Motoren durch deren geringen Preis fragwürdig erscheint, rechtfertigt sich auch bei ihnen ein derartiger Aufwand, da der nachträgliche Ausbau

¹Vgl. Chalmers *et al.* (1988), Fischer (1995) sowie Moczala *et al.* (1998) für eine detaillierte Beschreibung dieser Motoren.

²Alle Angaben gemäß ZVEI (Zentralverband der elektrotechnischen Industrie).

eines fehlerhaften Motors aus einem Kraftfahrzeug oder einem Haushaltsgerät deutlich teurer wäre. Die Prüfung muß aber wegen des geringen Wertes des Einzelstücks kostengünstig und in Hinblick auf die hohe Produktionszahl und die hohe Produktionsgeschwindigkeit mit Taktzeiten im Sekundenbereich sehr schnell erfolgen (vgl. Filbert (2003)).

Warum akustische Güteprüfung?

Die häufigsten Fehler an kleinen Kommutatormaschinen treten, wie von O'Donnell (1985) sowie O'Donnell (1987) gezeigt, im mechanischen Teil auf. Besonders oft treten Fehler an Lagern und am Kommutator auf. Obwohl andere Größen (wie die elektrischen) viel einfacher und mit weniger Fehlern zu messen sind, ist darum eine akustische Güteprüfung erforderlich, weil sich die mechanischen Fehler im Schallsignal deutlich stärker bemerkbar machen als z.B. im Stromsignal.

Die wichtigsten Ursachen für eine Geräuscentwicklung in Kleinmotoren sind nach Moczala *et al.* (1998):

- Periodische Gehäuseverschiebung
Ursache sind Pendelbewegungen des Ankers durch das ungleichmäßige Drehfeld oder periodische Belastungen.
- Magnetisch verursachte Geräusche
Diese Geräusche entstehen durch Wirbelkräfte, die Körperschall in Läufer und Ständer erzeugen, sowie durch magnetostriktive Dehnungen, die periodische Ständerverformungen hervorrufen.
- Aerodynamische Geräusche aus turbulenten Luftströmungen
Der aerodynamische Geräuschanteil besteht aus einem Grundrauschen, das durch regellose Wirbelablagerungen der zur Kühlung benötigten Umgebungsluft entsteht.
- „Unwuchtgeräusche“
Unwuchten verursachen Schwingungen des Ankers, die über das Gehäuse abgegeben werden. Da Unwuchten unerwünscht sind, sind diese Geräusche Ausdruck eines Fehler.
- Bürstengeräusche
Bürstengeräusche entstehen durch periodische Stöße von Kontaktlamellen und Kohlebürste. Bei einem außergewöhnlichen Auftreten deuten Bürstengeräusche auf einen der häufig auftretenden Fehler an den Kommutatoren hin.
- Getriebegeräusche
Die Schwingungsanregung erfolgt mit den Zahneingriffen. Außergewöhnliche Getriebegeräusche sind Ausdruck eines Fehlers im Getriebe.

- Lagergeräusche
Ursache von Lagergeräuschen sind das Abrollen im Lager selbst und Unwuchtkräfte des Läufers. Die relativ häufigen Lagerfehler verursachen ein verändertes Lagergeräusch.

Die letzten vier Geräusche können bei einem starken oder außergewöhnlichen Auftreten auf den Fehler hindeuten, der obiger Übersicht genannt wird. Im weiteren Verlauf der Arbeit (besonders in Kap. 8.2.2) werden die Fehler mit den angegebenen Großbuchstaben nach folgendem System referenziert:

- Fehlerfreie Motoren: „Fehlerklasse“ A,
- außergewöhnliche Geräusche, die keiner der anderen Gruppen zugeordnet werden können (allgemeiner Fehler): Fehlerklasse B,
- Unwuchtgeräusche: Fehlerklasse C,
- Bürstengeräusche: Fehlerklasse D,
- Getriebegeräusche: Fehlerklasse E,
- Lagergeräusche: Fehlerklasse F.

Ebenso sind viele Montagefehler von Waschmaschinen, z.B. schlecht montierte Teile, akustisch wirksam. Sie verursachen beim Schleudern oder anderen Programmen erhebliche Geräusche, finden aber nur einen marginalen Niederschlag in anderen Größen.

Warum objektive akustische Güteprüfung?

Derzeit findet die akustische Güteprüfung in den beiden Untersuchungsbereichen fast vollständig manuell statt. Die Motivation für eine automatisierte Prüfung liegt nicht nur im Zwang nach gesenkten Personalkosten, sondern auch in einer objektivierten Güteprüfung. Nach Feldmann und Keiper (1994) hängt die Zuverlässigkeit subjektiver Prüferinnen und Prüfer stark von der Tagesform der einzelnen Prüferin sowie den unterschiedlichen, subjektiven Prüfkriterien der einzelnen Personen ab. Fausten (2002) untersuchte die Ergebnisse subjektiver Prüfer noch detaillierter und kam zu dem Schluß, daß auch ausgeruhte PrüferInnen nur eine Erkennungsrate von 80% erreichen, während sie bei bei Schichtende auf 68% sinken kann.

Dennoch ist das automatisierte Prüfsystem dem menschlichen Prüfer hinsichtlich seiner Abstraktions- und Assoziationsfähigkeiten deutlich unterlegen. Ein automatisches Prüfsystem muß bei jeder kleinen Typenänderung mit einer großen Stichprobe trainiert werden. Weiterhin sind aufgrund des geringen Stückpreises insbesondere der Motoren hohe Fertigungstoleranzen zugelassen. Diese führen zu einer großen Streuung der akustischen Merkmale innerhalb der Gut-Klasse selbst eines Typs, die das Prüfsystem abstrahieren muß. Demgegenüber kann der menschliche Prüfer bereits aus Paarvergleichen einiger weniger Muster charakteristische Eigenschaften der

Geräusche erlernen und in vielen Fällen sogar auf verschiedene andere Typen übertragen. In den letzten Jahren wurden jedoch Klassifikationsverfahren vorgeschlagen, die solche Schwankungen akzeptieren und mit einer geringen Stichprobe auskommen (z.B. Fausten (2002)).

Verfahren der objektiven akustischen Güteprüfung

Automatische Prüfsysteme werten meßbare Signale der Geräte aus, hier den abgestrahlten Schall, um zu einer Diagnoseentscheidung zu kommen. Man kann sie nach Beilharz *et al.* (1995) in zwei Gruppen unterteilen: Modellbasierte Verfahren und merkmalsbasierte Verfahren. Erstere schätzen zum einen die aktuellen Parameter der Maschine (Parameteridentifikation, Beilharz (1999)) über einen Least-squares-Ansatz. Bei Fehlern ergeben sich Abweichungen zu Nominalparametern. Zum anderen generieren sie aus einer fehlerfreien Stichprobe ein Modell der Maschine. Die Differenz zwischen dem modellbasiert geschätzten und dem gemessenen Signal einer unbekannt Maschine gibt Auskunft über den Fehlerzustand (Gühmann (1995), Röpke (1997)). Die Nachteile dieser Verfahren bestehen darin, daß das akustische Modell einer kompletten Maschine in den meisten Fällen unbekannt (wie bei Waschmaschinen) oder nichtlinear (wie bei Kleinmotoren, vgl. Fausten (2002)) ist.

Die zweite Gruppe, die darum bei Schallsignalen angewendet wird, wertet die Signale direkt aus, wie in Abb. 2.1. Dazu werden aus den Signalen Merkmale extrahiert und mit ihnen eine Fehlerklassifikation durchgeführt (Barschdorff und Becker (1991)).

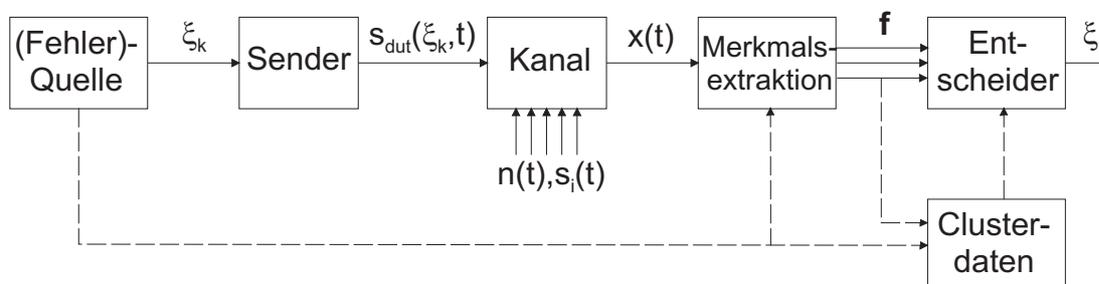


Abb. 2.1: Struktur eines merkmalsbasierten Fehlerdiagnosesystems

Für die Fehlerklassifikation wird angenommen, daß das akustische System der Maschine bei Vorliegen des Fehlers ξ_k das Schallsignal $s_{\text{dut}}(\xi_k, t)$ erzeugt, das durch ein Mikrophon aufgenommen wird. Die Entscheidung, um welchen Fehlerzustand es sich handelt, erfolgt durch Zuordnung des Signals zu einem Cluster, das einen Fehler repräsentiert und zuvor in einem überwachten oder nicht-overwachten Lernvorgang aus einer Lernstichprobe gewonnen wurde. Als Klassifikatoren werden sowohl statistische und Fuzzy-Klassifikatoren als auch neuronale Netze eingesetzt (vgl. Parsons (1987), Niemann (1983), Ruske (1988) zu Klassifikationssystemen).

Die Zuordnung erfolgt über L Merkmale, die aus dem Observationssignal extrahiert und in einem Merkmalsvektor \mathbf{f} zusammengefaßt werden. Ein solches Merkmal

ist nur schwer definierbar und kann nach Parsons (1987) jede aus Meßgrößen abgeleitete Größe sein, die sich für die Erkennung eines Muster als hilfreich erwiesen hat. Die Auswahl aussagekräftiger und robuster Merkmale ist die schwierigste Aufgabe beim Entwurf eines Klassifikators. Sie kann nach zwei Prinzipien erfolgen, nach dem heuristischen unter Zugrundelegen der Erfahrung oder dem analytischen durch systematische Auswahl der besten Merkmale aufgrund bestimmter Gütekriterien. In der Praxis findet meist eine Kombination aus beiden Verfahren statt, indem heuristisch gefundene Merkmale auf ihre Unterscheidungsfähigkeit hin überprüft werden. Typische Merkmale in der akustischen Fehlerdiagnose werden u.a. von Stepper (1999), Jahnke (2001) und Fausten (2002) vorgeschlagen. Einen allgemeinen Überblick geben auch Murray und Penman (1997).

Vorverarbeitung für die akustische Güteprüfung

In einer Fertigungshalle existieren neben den Signalen der zu testenden Maschinen zahlreiche andere Geräusche, mit denen das gemessene Signal (Observationssignal) gestört ist. Sie können erhebliche Pegel annehmen und den des Nutzsignals erreichen oder gar übertreffen. Weiterhin ist das Signalgemisch durch den Nachhall verzerrt. Das heißt, das Observationssignal $x(t)$ entsteht aus dem fehlerrelevanten Maschinensignal $s_{\text{dut}}(\xi_k, t)$ durch die Überlagerungen mit anderen Signalen $s_i(t)$ und Rauschen $n(t)$ und ist mit der Raumimpulsantwort h gefaltet.

Diese Effekte beeinträchtigen die Klassifikation oder machen sie gar unmöglich. Typischerweise wird das Problem heutzutage dadurch gelöst, daß der Meßplatz durch eine schalldämmenden Kabine von der Produktion isoliert ist. Die Isolierung erfolgt hauptsächlich durch schalldämmendes Material (Stein- oder Glaswolle o.ä.) hinreichender Dicke. Die wesentlichen Nachteile einer solchen mechanischen Dämmung liegen zum einen in den hohen Kosten: Die Anschaffungskosten einer Kabine liegen bei ca. 60.000 bis 100.000 EUR. Zum anderen stellen sie eine erhebliche Beeinträchtigung des Produktionsprozesses dar. Sie benötigen durch ihr Volumen zusätzlichen Platz, so daß die Geräte aus der Linie ausgeschleust werden müssen. Während einer Messung müssen die Zugänge verschlossen sein, so daß zusätzliche Mechanik nötig ist und knappe Taktzeit verbraucht wird.

Der Einsatz der Signalverarbeitung zur Signalrekonstruktion und Störunterdrückung in der Vorverarbeitung kann eine solche Kabine überflüssig machen oder wenigstens die Anforderungen an die Dämmeigenschaften reduzieren und damit Kosten sparen.

In der Literatur werden zwei Methoden vorgeschlagen. Die erste Methode ist, möglichst robuste Merkmale zu entwickeln, die unempfindlich gegen äußere Einflüsse sind. Eine Übersicht über robuste Merkmale wurden u.a. von Murray und Penman (1997) angegeben. Aiordachioaie *et al.* (1997) benutzen ein Hopfield-Netz³, dessen Cluster sie mit den Fehlerklassen vergleichen.

³Rückgekoppeltes neuronales Netz, das durch das Verhalten von Spingläsern motiviert ist, vgl. Rojas (1996).

Da es durch die große Streuung innerhalb der Klassen ohnehin nicht einfach ist, für die hier betrachteten Untersuchungsobjekte signifikante Merkmale zu finden, ist die Suche nach Merkmalen, die darüber hinaus noch unempfindlich gegen Störungen sind, sehr schwierig. Darum wird diese Methode nicht weiter untersucht.

Die zweite Methode ist, zunächst das Signal der Maschine aus dem Observationsignal zu rekonstruieren und die Merkmale aus diesem zu berechnen. Sie wird seltener verwendet, denn die Rekonstruktion des Maschinensignals der Länge O (in Samples) bedeutet die Schätzung eines O -dimensionalen Parametervektors. Da die Länge des Signals deutlich (wenigstens eine Größenordnung) größer ist als die Anzahl der Merkmale für die Klassifikation⁴, hat das Rekonstruktionsproblem eine wesentlich höhere Dimension als die Extraktion. Beispiele für die (einkanalige) Signalrekonstruktion werden Lee und White (1996) und Griffin *et al.* (1991) angeführt. Lee und White (1996) rekonstruieren einen fehlerrelevanten impulshaltigen Signalanteil, wenn dieser durch eine Sinusschwingung und breitbandiges Rauschen überlagert wird. In einem zweistufigen Verfahren werden beide überlagerten Anteile mit einem LMS-Algorithmus⁵ und einem QR-LSL-Algorithmus⁶ herausgefiltert. Griffin *et al.* (1991) arbeiten mit den Geräuschen von Raketenantrieben, die sie durch die Wagner-Ville-Verteilungen aus der Geräuschmischung während eines Raketenstartes extrahieren.

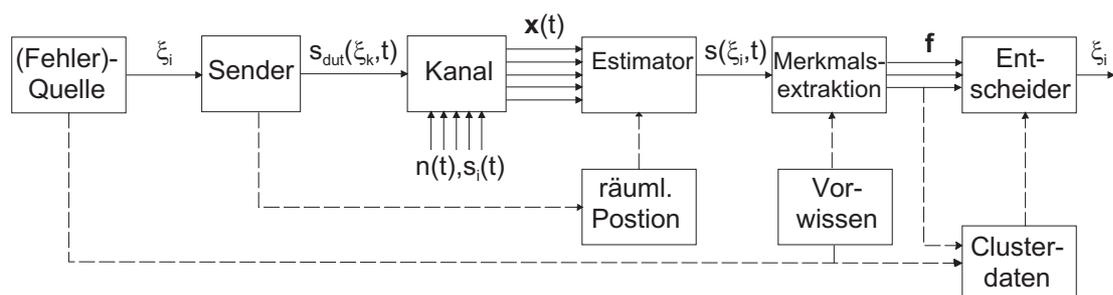


Abb. 2.2: Struktur des vorgeschlagenen Fehlerdiagnosesystems mit einer Erweiterung um einen mehrkanaligen Signalschätzer

In dieser Arbeit wird vorgeschlagen, aus einer mehrkanaligen Signalaufnahme das Maschinensignal zu schätzen, wie es in Abb. 2.2 dargestellt ist. Dabei wird nicht nur ein Observationssignal, sondern es werden N Observationen aufgenommen. Damit erhält man zusätzliche Informationen über das Motorsignal und die Störungen. Sie ist die Grundlage für die verbesserte Störunterdrückung und Signalrekonstruktion, die im folgenden beschrieben wird.

⁴ Anderenfalls wäre die Merkmalsextraktion sinnlos.

⁵ Least-Mean-square-Algorithmus, vgl. Haykin (1996).

⁶ QR-decomposition-based least-squares lattice, vgl. Haykin (1996).

Kapitel 3

Mehrkanalige Schallmessung

Mikrofon-Arrays sind ausgedehnte Schallsensoren. Um die folgenden wichtigen Bedingungen für die blinde Quellentrennung und die adaptiven Beamformer herleiten zu können, werden in diesem Kapitel kurz die Grundlagen der Schallausbreitung (Kap. 3.1) und der ausgedehnten Sensoren (Kap. 3.2) gelegt.

3.1 Ausbreitung der Schallsignale im Raum

3.1.1 Entstehung und Ausbreitung von Schallwellen

Wenn Teile eines elastischen Mediums zu Vibrationen angeregt werden, kommt es zu einer Ausbreitung dieser Schwingungen in Form von Druckwellen (Günther *et al.* (1994)). Die mathematische Beschreibung eines solchen Wellenfeldes in der Luft erfolgt über den Druck p und die Schallschnelle v , die die Geschwindigkeit angibt, mit der sich die Luftteilchen infolge der Druckänderung bewegen. Beide Größen sind über das Newtonsche Gesetz und die Kontinuitätsgleichung verknüpft (Heckl und Müller (1994)). Aus dem Newtonschen Gesetz folgt in erster Näherung

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial v_x}{\partial t} \quad -\frac{\partial p}{\partial y} = \rho \frac{\partial v_y}{\partial t} \quad -\frac{\partial p}{\partial z} = \rho \frac{\partial v_z}{\partial t}$$

mit den Raumkoordinaten x, y, z sowie der Zeit t . Die Kontinuitätsgleichung lautet für eine adiabatische Kompression mit der Kompressibilität der Luft K (adiabatischer Exponent) in erster Näherung

$$\frac{\partial p}{\partial t} = K \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \right).$$

Differenziert man diese Gleichungen erneut nach x, y und z bzw. der Zeit t , erhält man eine Wellengleichung zweiten Grades:

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = \frac{1}{K \rho} \frac{\partial^2 p}{\partial t^2}.$$

Eine analoge Gleichung kann man für die Schallschnelle angeben. Darum wird als Basisgröße für das Wellenfeld zur Vereinheitlichung die Variable s benutzt. Mit $c^2 = 1/K\rho$, der Proportionalitätskonstante zwischen den Energiespeichern, ergibt sich

$$\frac{\partial^2 s}{\partial x^2} + \frac{\partial^2 s}{\partial y^2} + \frac{\partial^2 s}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 s}{\partial t^2}. \quad (3.1)$$

Diese Differentialgleichung ist die allgemeine Differentialgleichung zur Beschreibung von Schallfeldern und Wellen. Der einzige Parameter c ist die Ausbreitungsgeschwindigkeit des Feldes.

Der Ansatz zur Lösung in kartesischen Koordinaten lautet allgemein nach Johnson und Dudgeon (1993):

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

Konkret wird ein exponentieller Ansatz verwendet. ω beschreibt die Periodizität in der Zeit und k_x, k_y, k_z die Periodizität im Raum:

$$s(x, y, z, t) = A \cdot e^{j(\omega t - k_x x - k_y y - k_z z)}. \quad (3.2)$$

Nach Einsetzen in (3.1) erhält man

$$k_x^2 s + k_y^2 s + k_z^2 s = \frac{\omega^2 \cdot s}{c^2}$$

und es ergibt sich die Bedingung

$$k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c^2}. \quad (3.3)$$

Die Lösung in (3.1) wird als monochromatische (nur mit einer Frequenz schwingende), planare Welle bezeichnet. Planar bedeutet, daß der Wellenzahlvektor $\vec{k} = [k_x k_y k_z]^1$ keine Funktion der Zeit ist, was äquivalent mit einer ebenen Wellenfront ist, wie man im folgenden sehen wird. In Ortsvektor-Schreibweise lautet (3.2)

$$s(\vec{x}, t) = A \cdot e^{j(\omega t - \vec{k}\vec{x})}. \quad (3.4)$$

Da t und \vec{x} durch $\omega t - \vec{k}\vec{x}$ miteinander verbunden sind, handelt es sich bei dieser Lösung um ein sich ausbreitendes Schallfeld. Am Ort $\vec{x} + \Delta\vec{x}$ ist der Druckzustand zum Zeitpunkt $t + \Delta t$ erreicht, den der Ort \vec{x} zum Zeitpunkt t hatte. Denn mit $\vec{\xi}$, der Richtung des Vektors \vec{k} , $\Delta x = c \cdot \Delta t$ und (3.3) gilt:

$$\omega t - \vec{k}(\vec{x} + \Delta\vec{x}) = \omega t - \vec{k}\vec{x} + \frac{\omega}{c} \vec{\xi} \Delta\vec{x} \quad (3.5)$$

$$= \omega t + \omega \Delta t - \vec{k}\vec{x}. \quad (3.6)$$

¹Generell werden in dieser Arbeit Ortsvektoren (dreidimensionale Vektoren mit den Komponenten x, y, z und einem Bezug physikalischen Raum) mit dem Symbol $\vec{\bullet}$ bezeichnet. Allgemeine N -dimensionale Vektoren werden hingegen mit fettgedruckten Variablen gekennzeichnet.

Der Vektor \vec{k} zeigt damit in die Richtung der schnellsten Änderung des Wellenfeldes und somit in die Ausbreitungsrichtung der Welle. Wenn $\Delta\vec{x} \perp \vec{k}$ ist, so ist $\Delta t = 0$ und der Druckzustand ist bereits zum gleichen Zeitpunkt gleich. Die Orte gleichen Druckes befinden sich auf einer Ebene mit dem Normalenvektor \vec{k} und werden als Wellenfront bezeichnet.

Obwohl \vec{k} und ω für die Charakterisierung der Welle ausreichend sind, werden hier alle Größen mit ihren Einheiten und den wichtigsten Gleichungen aufgeführt, die für die Beschreibung der Welle allgemein benutzt werden und auch im folgenden Verwendung finden:

- Ausbreitungsgeschwindigkeit $[c] = \frac{m}{s}$ $c = \frac{\omega}{|\vec{k}|} = \frac{\lambda}{T}$
- zeitliche Kreisfrequenz $[\omega] = \frac{1}{s}$
 $\omega = 2\pi f$
- Wellenzahl, räumliche Kreisfrequenz (beschreibt ebenfalls die Ausbreitungsrichtung) $[\vec{k}] = \frac{1}{m}$
- Wellenlänge $[\lambda] = m$
 $\lambda = \frac{2\pi}{|\vec{k}|}$ $\vec{k} = \frac{\omega\vec{\xi}}{c}$
- Ausbreitungsrichtung der Welle, Richtung der Wellennormalen (ohne Einheit)
 $\vec{\xi} = [\xi_x \xi_y \xi_z]^T$ $|\vec{\xi}| = 1$ $\vec{k} = |\vec{k}| \cdot \vec{\xi}$
Häufig wird $\vec{\xi}$ mit den Winkeln θ und ϕ bezeichnet, wofür die Bezeichnungen Alzimuth und Elevation üblich sind.

Alternativ zu (3.1) kann die Differentialgleichung des Schallfeldes auch in Kugelkoordinaten aufgestellt und gelöst werden. Sie lautet mit $\vec{r} = [r, \phi, \theta]$

$$\frac{\partial^2(sr)}{\partial r^2} = \frac{1}{c^2} \frac{\partial^2(sr)}{\partial t^2}, \quad (3.7)$$

wenn aus den Randbedingungen ersichtlich ist, daß s nicht von ϕ und θ abhängt, also ein kugelsymmetrisches Problem vorliegt. Solche Bedingungen liegen im Nahfeld einer punktförmigen Quelle vor. Die monochromatische Lösung lautet nun:

$$s(r, \phi, \theta, t) = s(\vec{r}, t) = A \cdot e^{j(\omega t - kr)}. \quad (3.8)$$

Der Vektor \vec{k} hat nur eine Komponente in \vec{r} -Richtung und die Wellenfront ist kugelförmig. Daher wird eine Welle im Nahfeld auch als Kugelwelle bezeichnet.

Der Schall ist für die Informationsübertragung geeignet, denn jedes Signal erfüllt die Wellengleichung. Dadurch bleibt die Form der Welle erhalten, wenn sie sich durch den Raum ausbreitet, und man erhält an beliebigen Orten nach einer Zeitverzögerung ein Abbild der Welle. In der Realität treten weitere Effekte auf, die die Welle zusätzlich beeinflussen und möglicherweise korrigiert werden müssen. Solche Effekte sind Reflexion, Beugung, Brechung, Dispersion und Dämpfung.

Die Dispersion wird vernachlässigt, weil in den geringen Abständen der Versuchsaufbauten die Medien als dispersionsfrei angenommen werden. Die Dämpfung gilt als konstant und wird ebenfalls vernachlässigt. Weiterhin wird keine Beugung betrachtet.

Trifft eine Welle an eine Grenzfläche zwischen zwei Medien unterschiedlicher Dichte, so wird sie reflektiert und gebeugt. In der vorliegenden Anwendung (Luft in Fabrikhallen oder Laborräumen) tritt ein solcher Dichteunterschied nur an den Grenzflächen der Objekte auf, wenn die Luft als homogen angenommen wird. Da die Welle nicht in den Objekten oder Raumbegrenzungen aufgenommen werden soll, ist nur die Reflexion von Bedeutung. Ihr für diese Arbeit wichtigstes Gesetz ist: $\vec{k}_i \vec{x} = \vec{k}_r \vec{x}$ und damit $|k_i| \sin(\theta_i) = |k_r| \sin(\theta_r)$ und $\theta_i = \theta_r$.

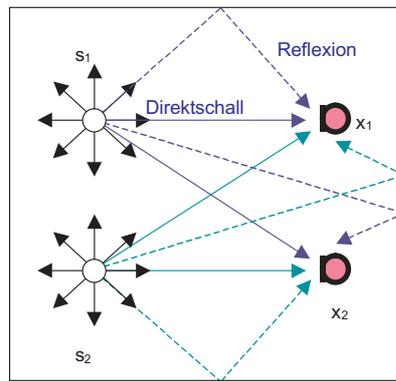


Abb. 3.1: „Mehrwegeüberlagerung“ durch Reflexionen an den Wänden

Die Konsequenzen der Reflexionen sind, daß ein Signal über mehrere Wege zum Mikrofon gelangen kann. Das Mikrofon empfängt eine „Mehrwegeüberlagerung“ des Quellensignals, wie in Abb. 3.1 dargestellt. Da die Pfade eine unterschiedliche Länge haben und damit die Laufzeit variiert, erreichen jedes Mikrofon mehrere zeitverzögerte Kopien des Signals. Dieses Verhalten kann durch einen räumlichen Filter (zur Erklärung eines solchen Filters siehe Kap. 3.1.2) beschrieben werden.

$$s_x(t, \vec{x}) = \int_{\Delta_{max}}^{\Delta_{max}} A(\tau) s_0(t - \tau) d\tau \quad (3.9)$$

Die Impulsantwort eines solchen Filters wird Raum-Impulsantwort genannt. Eine typische Raumimpulsantwort ist in Abb. 3.2 dargestellt. Darüberhinaus entsteht durch viele Reflexionen kohärenter und gerichteter Wellen ein diffuses Schallfeld.

3.1.2 Wellenbeschreibung im Frequenz-Wellenzahlbereich

Die Beschreibung von Wellenfeldern (räumlich-zeitlichen Signalen) im Wellenzahl-Frequenzbereich ist ähnlich bedeutsam wie die Beschreibung von zeitlichen Signalen im Frequenzbereich. Die Transformation in diesen Bereich erfolgt durch eine vierdimensionale Fouriertransformation.

$$S(\vec{k}, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(\vec{x}, t) \cdot e^{-j(\omega t - \vec{k} \vec{x})} d\vec{x} dt \quad (3.10)$$

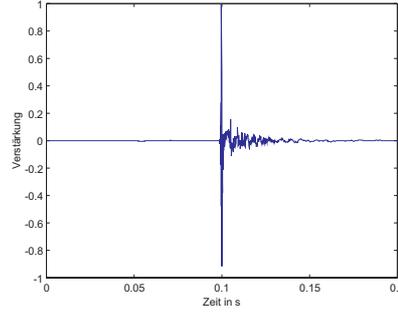


Abb. 3.2: Typische Impulsantwort eines kleinen Raumes mit einer Nachhallzeit von 150 ms

Beliebige Wellenfelder können mit der inversen Fouriertransformation aus diesen Wellen zusammengefaßt werden:

$$s(\vec{x}, t) = \frac{1}{(2\pi)^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(\vec{k}, \omega) \cdot e^{j(\omega t - \vec{k}\vec{x})} d\vec{x} dt. \quad (3.11)$$

Die Transformierte einer planaren, monochromatischen Welle ist

$$\begin{aligned} S_{\text{planar}}(\vec{k}, \omega) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A \cdot e^{-j(\omega - \omega_0)t + j(\vec{k} - \vec{k}_0)\vec{x}} d\vec{x} dt \\ &= A \delta(\vec{k} - \vec{k}_0) \delta(\omega - \omega_0), \end{aligned}$$

wobei δ die Deltafunktion und $\delta(\vec{k})$ die verkürzte Schreibweise für die Deltafunktion des Vektors \vec{k} ist.

Weitere wichtige Unterräume des Wellenzahl-Frequenzbereichs sind die breitbandigen propagierenden und die isotropen Felder. Jede propagierende Welle hat nur Werte ungleich null entlang der Linie

$$\vec{k} - \frac{\omega \vec{\xi}}{c} = 0 \quad c\vec{k} = \omega \vec{\xi}. \quad (3.12)$$

Mit jeder zeitlichen Frequenz ist durch Schallgeschwindigkeit c und die Ausbreitungsrichtung $\vec{\xi}$ eine Wellenzahl verbunden

$$S(\vec{k}, \omega) = S(\omega) \cdot \delta\left(\vec{k} - \frac{\omega \vec{\xi}}{c}\right).$$

Ein isotropes Wellenfeld breitet sich in alle Richtungen gleich aus. Es hat einen kegelförmigen Unterraum mit $c \cdot |k| = \omega$. Im Vergleich mit (3.12) erkennt man, daß die Richtungsabhängigkeit fehlt. Isotrope Wellenfelder werden oft für die Beschreibung diffuser Wellenfelder, die durch zahlreiche Reflexionen gerichteter Wellenfelder entstehen, genutzt. Wenn die diffusen Wellenfelder stationär sind, sind sie auch isotrop (Johnson und Dudgeon (1993)).

In Analogie zur Beschreibung zeitlicher Signale kann eine Filterung definiert werden. Dazu ist es notwendig, ein spatio-temporales Filter über eine vierdimensionale Faltung (räumliche und zeitliche Faltung) zu erklären. Diese Faltung ist durch

$$y(\vec{x}, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\vec{x} - \vec{\chi}, t - \tau) \cdot s(\vec{\chi}, \tau) d\vec{\chi} d\tau \quad (3.13)$$

definiert.

3.2 Einführung in Mikrofon-Arrays

Ein Mikrofon-Array ist ein diskreter, räumlich ausgedehnter Schallaufnehmer. Benutzt man ein Array wie in dieser Arbeit dazu, das Signal aus einer bestimmten Richtung $\vec{\xi}$ zu schätzen, bedeutet das die Schätzung einer bestimmten Wellenzahl, d.h. einer räumlichen Frequenz. Man kann die Aufgabe also als eine spektrale Schätzung ansehen (Steinberg (1976)). Das heißt, daß die Methoden der eindimensionalen spektralen Schätzung mit einigen Umdefinitionen auf Mikrofon-Arrays angewendet werden können².

Die spektrale Schätzung aus der diskreten Fouriertransformation führt auf einen Delay-&-Sum-Beamformer. Sog. „High resolution techniques“³, wie die Minimum-Varianz-Schätzung, führen auf den Minimum-Varianz- und den Maximum-Likelihood-Beamformer.

Dieser Abschnitt beschäftigt sich zunächst mit den mathematischen Konsequenzen einer räumlich ausgedehnte, jedoch begrenzten (3.2.1) und räumlichen diskreten (3.2.2) Signalaufnahme. Die Varianten der Fokussierung durch unterschiedliche Beamformer ist Inhalt des nächsten Kapitels (Kap 4).

3.2.1 Räumlich ausgedehnte Signalaufnahme

Die Welle wird mit einem Mikrofon-Array in elektrische Signale umgewandelt. An den Orten \vec{x}_m befinden sich dazu M Mikrofone mit den elektrischen Ausgangssignalen $y(t)$. Die Ausgangssignale sind eine lineare Funktion von $s(\vec{x}_m, t)$, wenn ideale Sensoren (d.h. nicht bandbegrenzt und linear) vorausgesetzt werden. In diesem Abschnitt soll die richtungselektive Wirkung gezeigt werden, wenn die gemessenen Signale (Observationen) zu einem Signal zusammengefaßt werden sollen.

Zur Vereinfachung soll zunächst angenommen werden, daß der Sensor nicht diskret ist, sondern an allen Punkten \vec{x} aufgenommen wird⁴. Mathematisch faßt man

²Man kann die Analogie zu eindimensionalen Signalen weiter ausbauen. Ein (omni-direktives) Einzelmikrofon wertet das Wellenfeld nur an einem Punkt aus. Das entspricht für eindimensionale Signale der Aufnahme zu einem Zeitpunkt, so daß keine spektrale Schätzung möglich ist. Ein Array wertet das Wellenfeld hingegen an diskreten Punkten in einem Bereich aus. Das Gegenstück ist eine zeit-diskrete Aufnahme in einem Fenster, so daß hier eine spektrale Schätzung möglich ist.

³So werden alle Verfahren genannt, die mit parametrischen und anderen Schätzverfahren eine höhere Auflösung als die Fouriertransformation erreichen.

⁴Anschaulich kann man sich das als Parabol-Antenne vorstellen.

dies dadurch, daß das Signal an jedem Ort ein bestimmtes Gewicht $w(\vec{x})$ bekommt und dieses Gewicht außerhalb der Sensorregion null ist (Shadingfunktion⁵). Eine typische Shadingfunktion ist in Abb. 3.3 dargestellt. Das Analogon zur Shadingfunktion ist die Fensterfunktion (u.a. Lüke (1995)) für zeitlichen Funktionen.

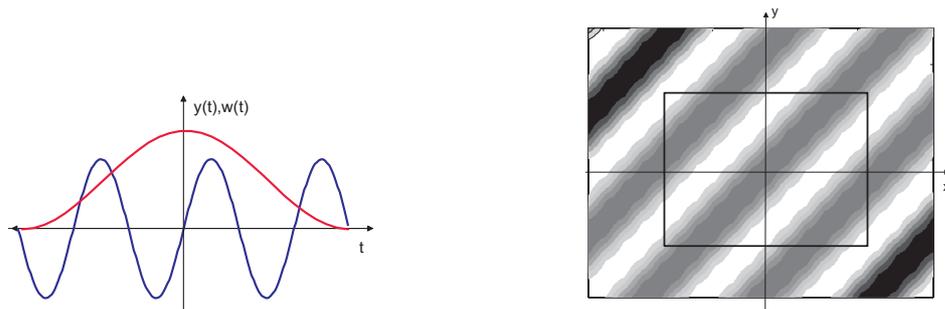
$$y(\vec{x}, t) = w(\vec{x}) \cdot s(\vec{x}, t)$$

Da hier eine spektrale Schätzung erfolgen soll, wird $y(\vec{x}, t)$ gemäß (3.10) in den Wellenzahl-Frequenz-Bereich transformiert.

$$Y(\vec{k}, \omega) = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} W(\vec{k} - \vec{l}) \cdot S(\vec{l}) d\vec{l} \quad (3.14)$$

Gleichung (3.14) beschreibt eine räumliche Faltung zwischen der Transformaten des Signals und der Transformaten der Shadingfunktion nach (3.15). Oft wird die Transformaten der Shadingfunktion als Arraypattern bezeichnet.

$$W(\vec{k}) = \int_{-\infty}^{\infty} w(\vec{x}) \cdot e^{j\vec{k}\vec{x}} d\vec{x} \quad (3.15)$$



(a) Sinusschwingung mit einem typischen Zeitfenster (Hanning-Fenster)

(b) Ausgedehntes rechteckiges Array in einem Wellenfeld einer Frequenz. Die Grautöne zeigen den Druck am jeweiligen Ort an

Abb. 3.3: Analogie zwischen Shadingfunktion und einem zeitlichen Fenster

Um die Wirkung der Faltung genauer zu untersuchen, sei das Wellenfeld eine Überlagerung planarer monochromatischer Wellen s_i gleicher Frequenz aus den Richtungen ξ_i :

$$y(\vec{x}, t) = w(\vec{x}) \cdot \sum_i A_i e^{j(\omega_0 t - \vec{k}_i \vec{x})} = \sum_i A_i e^{j\omega_0(t - \vec{\xi}_i \vec{x}/c)}$$

⁵Shading- und Aperturfunktion werden häufig für die Beschreibung eines ausgedehnten Sensors verwendet. Die Aperturfunktion wird eher benutzt, wenn der Einfluß der Sensorausdehnung untersucht werden soll. Ihre wichtigste Eigenschaft ist, für Orte außerhalb des Sensors den Wert Null anzunehmen. Mit der Shading-Funktion hingegen wird der Einfluß von unterschiedlichen Gewichten an jedem Ort innerhalb des Sensor untersucht. Da beide Begriffe fast synonym sind, wird im folgenden nur von der Shadingfunktion gesprochen.

Mit dem Array soll nun ein Signal s_n aus der Richtung $\vec{\xi}_n$ geschätzt werden. Die Transformierte jeder Welle wird gemäß (3.4) durch $S_i(\vec{k}_i, \omega) = A_i \delta(\omega - \omega_0) \delta(\vec{k} - \vec{k}_i)$ beschrieben. Mit der Ausblendeigenschaft der Deltafunktion wird (3.14) zu

$$\begin{aligned} Y(\vec{k}, \omega) &= \sum_i A_i \cdot \delta(\omega - \omega_0) W(\vec{k} - \vec{k}_i) = \sum_i A_i \cdot \delta(\omega - \omega_0) W\left(\frac{\omega_0}{c}(\vec{\xi} - \vec{\xi}_i)\right) \\ Y(\vec{\xi}, \omega_0) &= \sum_i A_i \cdot W\left(\frac{\omega_0}{c}(\vec{\xi} - \vec{\xi}_i)\right) \end{aligned} \quad (3.16)$$

Da also das Arraypattern die Selektivität bestimmt, soll es näher untersucht werden. Durch die Analogie zur Transformierten einer Fensterfunktion ist es eine si-Funktion, d.h. es ist eine periodische Funktion mit abklingender Amplitude. Im Fall eines Linienarrays mit konstanter Gewichtung w_x (Abb. 3.4) entspricht es exakt einer si-Funktion⁶.

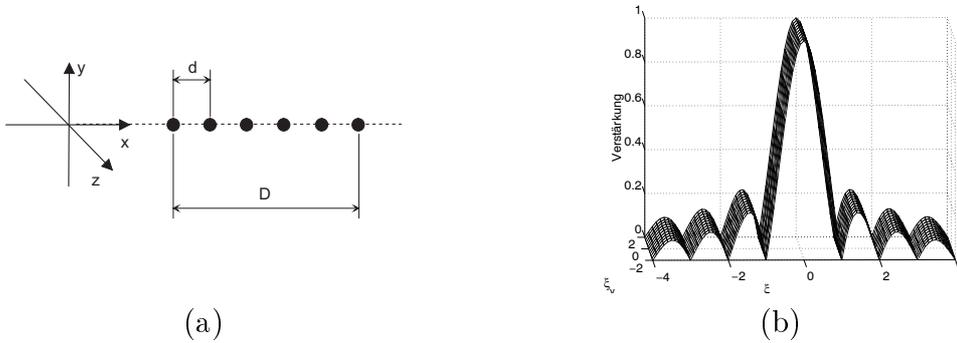


Abb. 3.4: Linienarray: Anordnung (a) und Arraypattern (b)

$$w(\vec{x}) = w_x(x) \delta(y) \delta(z) \quad w(x) = \begin{cases} 1 & : |x| \leq \frac{1}{2}D \\ 0 & : \text{sonst} \end{cases} \quad (3.17)$$

Mit (3.15) ergibt sich das Arraypattern:

$$\begin{aligned} W(\vec{k}) &= \int_{-\frac{D}{2}}^{\frac{D}{2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w_x(x) \delta(y) \delta(z) \cdot e^{j\vec{k}\vec{x}} \cdot dx dy dz \\ &= \int_{-\frac{D}{2}}^{\frac{D}{2}} 1 \cdot e^{k_x x} dx \\ &= \text{si}\left(\frac{D}{2} k_x\right). \end{aligned} \quad (3.18)$$

Abb. 3.4b stellt ein beispielhaftes Arraypattern für eine bestimmte Frequenz dar.

Damit wird aus (3.16)

$$Y(\vec{\xi}_n, \omega_0) = \sum_i A_i \cdot W\left(\frac{\omega_0}{c}(\vec{\xi} - \vec{\xi}_i)\right) = A_n + \sum_{i \neq n} A_i \cdot \text{si}\left(\frac{D}{2} \frac{\omega_0}{c} (\xi_{x,i} - \xi_{x,n})\right).$$

⁶Meist wird eine andere Fensterfunktion (z.B. Flat-top- oder ein von-Hann-Fenster) verwendet, die ein anderes Arraypattern ergibt. Die grundsätzlichen Eigenschaften sind jedoch gleich, weshalb sich dieses Kapitel auf die si-Funktion beschränkt.

Der zu schätzende spektrale Anteil bleibt erhalten, weil die si-Funktion bei $\xi = \xi_0$ den Wert eins hat. Jedoch wird er durch die anderen Signalanteile überlagert, weil die si-Funktion im zweiten Term für $\vec{k} \neq \vec{k}_0$ und damit $\vec{\xi} \neq \vec{\xi}_0$ nicht überall verschwindet.

Die maximale Dämpfung erfährt ein Signal, wenn die zugehörige si-Funktion eine Nullstelle bei $\vec{\xi}_n$ hat. Dann hat das Signal aus dieser Richtung keinen Einfluß. Der Abstand der ersten Nullstelle ξ_{n1x} wird als Rayleigh-Auflösung bezeichnet:

$$2\pi \frac{D}{c} \xi_{n1x} = \pi. \quad (3.19)$$

Da (3.19) frequenzabhängig ist, werden die nicht aus der Richtung ξ_0 stammenden Wellen nicht vollständig unterdrückt, sondern nur gefiltert, wenn sie breitbandig sind.

Die Höhe der Nebenkeulen (Maxima der Shading-Funktion für $\vec{\xi} \neq \mathbf{0}$) bestimmt die minimale Dämpfung der Signale. Durch die gezeigte Analogie zwischen Shadingfunktionen und zeitlichen Fenstern wird klar, daß die Höhe der Nebenkeulen durch eine andere Shadingfunktion verändert werden kann. Die Verbesserung der Dämpfung wird jedoch durch eine Verschlechterung der Rayleigh-Auflösung 'erkauft'. Für eine ausführliche Diskussion möglicher Fensterfunktion siehe Hoffmann (1998) oder Oppenheim und Schafer (1995).

Durch die Symmetrie im Array kann es zu Mehrdeutigkeiten kommen, die nicht aufgelöst werden können. Da (3.18) beispielsweise nur von ξ_x , d.h. der x-Komponente der Wellennormalen, abhängt, liefern alle Wellen mit beliebigen Richtungskomponenten ξ_y und ξ_z den gleichen Beitrag zum geschätzten Signal. In Abb. 3.4b erkennt man, daß die Shadingfunktion einen konstanten Verlauf in y-Richtung hat, d.h. keine Selektivität in dieser Richtung besteht.

Da die am Beispiel eines Linien-Arrays vorgestellten Eigenschaften analog für alle beliebigen Arrays gelten, läßt sich zusammenfassend sagen, daß durch die Auswertung ausgedehnter Aufnahmeanordnungen eine Richtcharakteristik erreicht werden kann. Bei der Aufnahme mit nur einem (omni-direktionalen) Mikrofon hingegen ist die Ausdehnung $D = 0$ und damit die Hauptkeule der si-Funktion ausgedehnt. Das Mikrofon hat keine bevorzugte Richtung. Damit stellen ausgedehnten Sensoren die Basis für jeden Beamformer und jede mehrkanalige Signalverarbeitung dar.

3.2.2 Räumliche diskrete Signalaufnahme

Das Wellenfeld wird nun mit einem ausgedehnten Sensor an diskreten Orten \vec{x}_m aufgenommen, d.h. es erfolgt eine räumliche Abtastung. Die diskrete Shadingfunktion ist damit $w_{diskret} = \sum_m \delta(\vec{x} - \vec{x}_m)w(\vec{x})$. Wenn die Mikrofone wie in Abb. 3.4 in

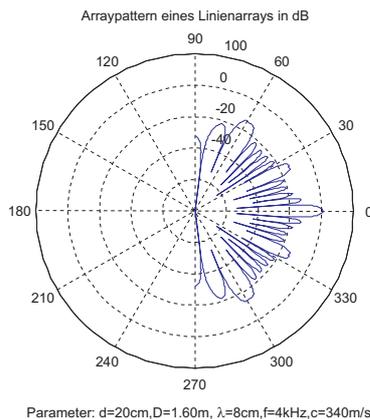


Abb. 3.5: Arraypattern eines diskreten Linienarrays in Abhängigkeit vom Einfallswinkel mit räumlichem Aliasing

einem Linienarray im Abstand d angeordnet sind, ergibt sich

$$\begin{aligned}
 W_{\text{diskret}}(\vec{k}) &= \int_{-\infty}^{\infty} \sum_m \delta(x - md) w(\vec{x}) \cdot e^{j\vec{k}\vec{x}} d\vec{x} \\
 &= \sum_m \delta\left(k_x - \frac{2\pi m}{d}\right) W(\vec{k}_x) \\
 &= \sum_m W\left(k_x - \frac{2\pi m}{d}\right) \tag{3.20}
 \end{aligned}$$

$$= \sum_m \text{si}\left(\frac{D}{2}\left(k_x - \frac{2\pi m}{d}\right)\right). \tag{3.21}$$

Dies entspricht dem Shannon-Theorem, nach dem die Abtastung auf ein kontinuierliches, jedoch bei $k_x = \frac{1}{d}$ wiederholtes Spektrum führt (vgl. Abb. 3.5). Das heißt, daß im Spektrum neue Mehrdeutigkeiten auftreten, da keine Bandbegrenzung für den Wellenzahlbereich vorgenommen werden kann⁷.

Dieser Effekt wird als räumliches Aliasing bezeichnet. Er ist unvermeidbar und muß auch für höhere Methoden der Signalverarbeitung beachtet werden, wenn sie auf einer spektralen Schätzung beruhen. Bei einer günstigen Wahl der Mikrofonabstände kann man das Aliasing für einen bestimmten Frequenzbereich verhindern. Damit begrenzt das Shannon-Theorem die Arraygröße für einen vorgegebenen Frequenzbereich und eine feste Mikrofonanzahl nach oben. Nach unten wird sie durch die Kohärenzlänge des diffusen Schalls begrenzt (vgl. Drews (1999)).

⁷Es gibt kein räumliches Anti-Aliasing-Filter, das man vor das Array setzen kann.

Kapitel 4

Beamforming

Mit Beamforming bezeichnet man die Ausnutzung der Eigenschaften eines Mikrofonarrays zur Fokussierung auf Schallwellen, die aus einer bestimmten Richtung kommen. Nach den Aussagen aus Kap. 3 bedeutet die Fokussierung die Schätzung eines räumlich-spektralen Anteils. Grundsätzlich lassen sich dazu alle aus der Signalverarbeitung bekannten Methoden der frequenzselektiven Schätzung anwenden.

Die wesentliche Aufteilung besteht zwischen adaptiven und konventionellen Beamformern. Konventionelle Beamformer haben eine feste Richtcharakteristik durch fest voreingestellte Parameter. Sie entsprechen den nicht-parametrischen Methoden der Signalverarbeitung (Vaseghi (1996)). Adaptive Beamformer¹ verändern ihre Richtcharakteristik in Abhängigkeit von den Signalen und ihrer Statistik mit Hilfe parametrischer oder stochastischer Methoden (Bayes-Methoden).

Allen Beamformern ist gemeinsam, daß sie auf zusätzliche Randbedingungen angewiesen sind. Dadurch unterscheiden sie sich von den in Kap. 5 vorgestellten blinden Verfahren der mehrkanaligen Signalverarbeitung. Meistens wird die Richtung benötigt, aus der das Nutzsignal kommt. Teilweise muß zusätzlich auch die Richtung der kohärenten Störgeräusche bekannt sein.

Zunächst wird in diesem Kapitel die Aufgabenstellung formuliert und mit den Erkenntnissen des letzten Kapitels formalisiert. Danach werden konventionelle und adaptive Beamformer eingeführt und analysiert. Die Analyse konzentriert sich auf die Probleme, die durch suboptimale Bedingungen auftreten, d.h. wenn die Voraussetzungen für die Ableitung der Bestimmungsgleichungen für die Koeffizienten nicht erfüllt sind.

¹Für einen adaptiven Beamformer ist es nicht notwendig, daß er nach jedem Abtastwert angepaßt wird, wie es bei adaptiven Filtern nötig wäre. Alle Beamformer, die auf parametrischen oder stochastischen Methoden basieren, werden ebenfalls adaptiv genannt, auch wenn die Adaption blockweise erfolgt.

4.1 Definitionen und Annahmen über die Signale

In dieser Arbeit wird der Beamformer für die Schätzung eines zeitlichen Verlaufs eines Nutzsignals $\hat{s}_{\text{Target}}(t, \vec{\xi}_1)$ eingesetzt. Das Nutzsignal wird durch andere gerichtete Quellen im Raum und diffusen (nicht-gerichteten, räumlich wie zeitlich weißen) Störschall n_n^b mit der Leistung σ_n gestört.

Die gerichteten M Quellen werden zu einem Signalvektor zusammengefaßt, wobei o.B.d.A. s_1 das Nutzsignal ist. Dabei werden die Schallfelder am Ort der Entstehung betrachtet, so daß die Ortsabhängigkeit des Schallfeldes nicht notiert werden muß. Da Schall nur über die Änderung des Luftdrucks definiert ist, werden die Signale als mittelwertfrei angesehen.

$$\mathbf{s}^b(t) = [s_{\text{Target}}^b(t), s_{i1}^b(t), \dots, s_{iM-1}^b(t)]^T \quad (4.1)$$

Der obere Index \bullet^b gibt an, daß es sich um breitbandige Signale handelt, die im allgemeinen keine monochromatischen Wellen erzeugen. Die Leistungen der Quellen werden mit $\sigma_{s_1}^2, \dots, \sigma_{s_M}^2$ beschrieben.

Das Symbol $\hat{\bullet}$ dient zur Unterscheidung zwischen dem Originalsignal und seiner Schätzung. Im weiteren Verlauf wird die Ausgangsgröße des Schätzers zur besseren Unterscheidbarkeit auch als y oder \mathbf{y} (falls mehrere Größen geschätzt werden sollen) bezeichnet.

Das Schallfeld wird durch ein Mikrofon-Array mit N Mikrofonen erfaßt, danach verstärkt, mit einem angemessenen Tiefpaß gefiltert, mit $f_{ab} = \frac{1}{T}$ abgetastet und schließlich quantisiert, so daß die Zahlenfolge $x^b(n) = x^b(nT)$ entsteht. Für die Vereinfachung der Schreibweise und die leichtere Zuordnung der Variablen wird jedoch weiterhin mit der „diskreten Zeit“ t gearbeitet, so daß von hier an $x^b(t)$ nur zu den Zeitpunkten $t = nT$ einen Wert ungleich null hat (wenn es nicht explizit anders gekennzeichnet ist). Auf die Implikationen dieser Diskretisierung soll hier nicht eingegangen werden. Sie sind in zahlreichen Lehrbüchern beschrieben, z.B. Oppenheim und Schaffer (1995) oder Stearns und Hush (1994). Die so entstandenen Signale x_m werden analog den Quellensignalen zu einem Observationsvektor $\mathbf{x}^b(t) = [x_1^b(t), \dots, x_N^b(t)]^T$ zusammengefaßt.

Zusätzlich zu den Veränderungen durch die Meßeinrichtung werden die Quellensignale im Raum modifiziert. Diese Modifikationen wurden bereits in Kap. 3.1.1 als spatio-temporale Filterung durch die Überlagerung mit den Reflexionssignalen beschrieben (vgl. (3.13) und (3.9)). Die vierdimensionale räumlich-zeitliche Faltung (3.13) läßt sich hier in eine gewöhnliche zeitliche Faltung und eine Summe überführen, weil nur von einer endlichen Anzahl von M Quellen und dem isotropen Hinter-

grundschall ausgegangen wird:

$$\begin{aligned}
 x_{sn}^b(\vec{x}, t) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h^b(\vec{x} - \vec{\chi}, t - \tau) \cdot s^b(\vec{\chi}, \tau) d\tau d\vec{\chi} + n^b(\vec{x} - \vec{\chi}) \quad (4.2) \\
 &= \sum_{m=1}^M \int_{-\infty}^{\infty} h_{nm}^b(t - \tau) \cdot s_m^b(\tau) d\tau + n_n^b \\
 &= \sum_{m=1}^M h_{nm}^b * s_m^b + n_n^b.
 \end{aligned}$$

Der diffuse Schall wird durch ein isotropes Schallfeld modelliert, das damit die in Kap. 3.1.2 definierten Bedingungen erfüllt. Damit ist es sowohl zu den Quellen statistisch unabhängig als auch räumlich weiß, d.h. zwischen den Kanälen unkorreliert ($E\{n_k^b n_l^b\} = 0$ für $k \neq l$). Das Rauschen kann zu einem Rauschvektor $\mathbf{n}^b(t) = [n_1^b(t), \dots, n_N^b(t)]^T$ zusammengefaßt werden. Die Rauschleistung sei σ_n^2 .

Nach der Diskretisierung kann die Faltung durch einen FIR-Filter bzw. durch eine diskrete Faltung beschrieben werden. Man kann die (ebenfalls diskretisiert gedachten) Quellen- und Observationssignale in Vektorschreibweise zusammenfassen, indem man die von Lambert (1996) eingeführten FIR-Filtermatrizen verwendet. Diese Algebra erweitert den klassischen Algebra-Begriff auf Matrizen, deren Elemente FIR-Filter bzw. Polynome (z-Transformierte der Filter) sind.

Wenn sich das Wellenfeld gemäß (3.11) als Überlagerung monochromatischer, d.h. schmalbandiger, Wellen darstellen läßt, können die Mikrofonsignale ebenfalls als Aufnahmen dieser Überlagerungen angesehen werden. Filtert man darum schmalbandige Komponenten mit asymmetrischen hinreichend schmalen Bandpaßfiltern heraus, sind die gefilterten Signale Aufnahmen monochromatischer Wellen und können als solche behandelt werden.

Schmalbandige asymmetrische Filter sind Tiefpaßfilter, die durch eine Multiplikation mit $e^{j2\pi \cdot f_0 t}$ auf die Mittenfrequenz f_0 verschoben sind. Das entspricht, wie in Oppenheim und Schafer (1995) ausführlich beschrieben, einer Kurzzeit-Fouriertransformation (engl. short-time-fourier-transform, STFT), wobei die Breite des Bandpasses durch die Länge und die Form der Fensterfunktion bestimmt wird.

$$x_m^f(f_0, t) = \sum_{\theta=1}^K x_m^b(t + \theta) \cdot k(t) \cdot e^{-j2\pi\theta \cdot f_0} \quad (4.3)$$

Für eine vereinfachte Notation wird der obere Index \bullet^f im folgenden meist weggelassen, wenn er nicht zur Unterscheidung von den breitbandigen Signalen nötig ist². D.h. in dieser Arbeit ist in der Regel von schmalbandigen Signalen die Rede, wenn sie ohne oberen Index verwendet werden.

Die diskrete Fouriertransformation ist eine äquidistante Abtastung der z-Transformierten von x_m^b auf dem Einheitskreis. Wenn die in Oppenheim und Schafer (1995)

²Gelegentlich wird der obere Index f auch genutzt, um unterschiedliche Frequenzbänder zu unterscheiden.

genannten Voraussetzungen eingehalten werden (Vermeidung einer zyklischen Faltung), kann man die Faltung in (4.2) durch eine Multiplikation der schmalbandigen Komponenten ersetzen,

$$x_n = h_{nm}s_m \quad (4.4)$$

was sich in eine Matrixschreibweise überführen läßt.

$$\mathbf{x} = \mathbf{H} \cdot \mathbf{s} + \mathbf{n} = \begin{bmatrix} | & & | \\ \mathbf{h}_1 & \dots & \mathbf{h}_M \\ | & & | \end{bmatrix} \cdot \mathbf{s} + \mathbf{n} \quad (4.5)$$

Sind die Komponenten hinreichend schmalbandig, so daß sie als monochromatische Wellen angesehen werden können, gelten die Betrachtungen, die für propagierende monochromatische Quellen in Kap. 3.1.1 gemacht wurden. Damit unterscheidet sich die monochromatische Schallwelle an der Stelle des Mikrofons im Freifeld nur durch eine komplexe Drehung mit $e^{j\omega\tau}$ vom Quellensignal, wie in (3.6) gezeigt wurde. Darin ist τ die Verzögerung, die die Welle im Medium erfährt.

Für eine planare Welle (siehe Abb. 4.1a) ist die Wellenzahl und damit auch die Ausbreitungsrichtung nach (3.4) für alle Mikrofon-Positionen \vec{x}_n gleich. Von planaren Wellen kann ausgegangen werden, wenn der Abstand zwischen der Quelle und den Mikrofonen groß gegenüber den Mikrofonabständen (Fernfeld) ist. Wie in Abb. 4.1a sichtbar, ergibt sich die Verzögerung einer Welle aus der Richtung $\vec{\xi}_m$ zu $\tau = \vec{\xi}_m \cdot (\vec{x}_n - \vec{x}_0)/c$ (mit \vec{x}_0 , dem Ortsvektor der Arraymittelpunkts). Damit kann man einen Verzögerungsvektor

$$\mathbf{h}_n(\vec{\xi}_m) = \begin{bmatrix} e^{j\omega \frac{\vec{\xi}_m}{c} (\vec{x}_n - \vec{x}_0)} \\ \vdots \\ e^{j\omega \frac{\vec{\xi}_m}{c} (\vec{x}_N - \vec{x}_0)} \end{bmatrix} = \begin{bmatrix} e^{j2\pi f \tau_{1m}} \\ \vdots \\ e^{j2\pi f \tau_{Nm}} \end{bmatrix} \quad (4.6)$$

für jede Quelle definieren, der die Verzögerungen zu jedem Mikrofon enthält und der nur von der Einfallrichtung des Signals und der Positionen der Mikrofone abhängig ist. \mathbf{h}_1 sei der Verzögerungsvektor des Nutzsignals, das aus der Richtung ξ_1 kommt.

Für Kugelwellen (Nahfeld, Abb. 4.1b) kann man nicht von einer gleichen Ausbreitungsrichtung für alle Mikrofone ausgehen. Die Laufzeitdifferenzen ergeben sich aus dem Abstand $\vec{r}_{0m} = \vec{x}_0 - \vec{x}_m$ zwischen Quelle m und dem Array-Mittelpunkt sowie aus dem Abstand zwischen Quelle und Mikrofon m $\vec{r}_{0n} = \vec{x}_0 + \vec{x}_n$ zu

$$\Delta t = \frac{|\vec{r}_{0n} - \vec{r}_{0m}|}{c}.$$

Damit ergibt sich der Verzögerungsvektor zu

$$\mathbf{h}_m(\vec{x}_m) = \begin{bmatrix} e^{j \frac{|\vec{r}_{01} - \vec{r}_{0m}|}{c}} \\ \vdots \\ e^{j \frac{|\vec{r}_{0N} - \vec{r}_{0m}|}{c}} \end{bmatrix} = \begin{bmatrix} e^{j2\pi f \tau_{1m}} \\ \vdots \\ e^{j2\pi f \tau_{Nm}} \end{bmatrix}. \quad (4.7)$$

Im Nahfeld kann man die „Ausbreitungsrichtung“ der gesuchten Welle am Ort des Mikrofon-Arrays eigentlich gar nicht angeben und demzufolge auch nicht von einem

Verzögerungsvektor $\mathbf{h}_m(\vec{\xi}_m)$ sprechen. Man muß stattdessen die genaue Position der Quelle als Argument angeben: $\mathbf{h}_m(\vec{x}_m)$. Um jedoch eine unhandliche Doppelnotation „ $\mathbf{h}_m(\vec{\xi}_m)$ bzw. $\mathbf{h}_m(\vec{x}_m)$ “ zu vermeiden, wird der Verzögerungsvektor in dieser Arbeit generell mit $\mathbf{h}_m(\vec{\xi}_m)$ referenziert.

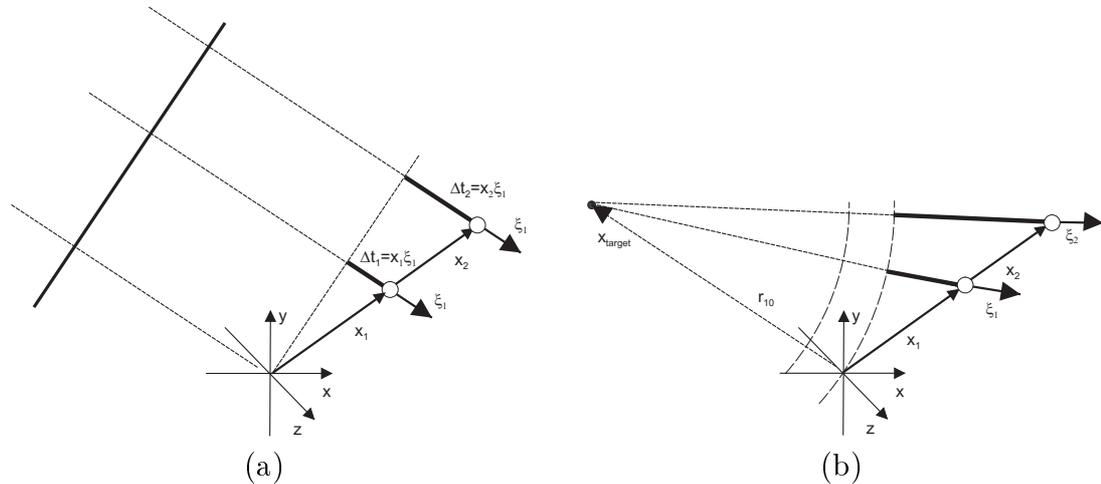


Abb. 4.1: Linienarray: im Fernfeld (a) und Nahfeld (b)

Herrschen keine Freifeldbedingungen, kann der tatsächliche Verzögerungsvektor des Nutzsignals nur exakt geschätzt werden, wenn die Raumimpulsantwort bekannt ist. Da dies nahezu nie der Fall ist, werden für seine Approximation ($\hat{\mathbf{h}}_1$) meistens Freifeldbedingungen angenommen. Dann benötigt man für seine Schätzung lediglich die Schallgeschwindigkeit c , die Einfallrichtung $\vec{\xi}_1$ bzw. die Position \vec{x}_1 des Nutzsignals und die Positionen der Mikrofone.

Während die Positionen der Mikrofone und die Schallgeschwindigkeit in der Regel als bekannt vorausgesetzt werden, müssen nur noch Einfallrichtung bzw. die Position des Nutzsignals ermittelt werden, was auf vielfältige Art geschieht. Aus den mehrkanaligen Verfahren selbst kann man sie mit sog. Tracking-Algorithmen oder einer DOA-Schätzung³ gewinnen (vgl. Johnson und Dudgeon (1993)). In letzter Zeit werden auch multimodale Ansätze verfolgt. In diesen erhält man die Informationen aus anderen Bereichen wie aus Bildern, Infrarotsensoren o.ä. In der Maschinendiagnose und in zahlreichen anderen Anwendungen, wie der Rekonstruktion des Sprechersignals im Auto, ist eine Schätzung jedoch gar nicht nötig, weil sämtliche geometrische Informationen a-priori bekannt sind.

Statistische Definitionen

Da die schmalbandigen Komponenten komplexwertig sind, muß der Formalismus der komplexwertigen linearen Algebra angewendet werden. Er basiert im wesentlichen darauf, daß das innere Produkt $c = \langle \mathbf{a} \mathbf{b} \rangle$ zweier Vektoren \mathbf{a} und \mathbf{b} anders als im reellen Fall definiert werden muß, um die Bedingung einzuhalten, daß das

³Direction of arrival, Einfallswinkel

innere Produkt eines Vektors mit sich selbst das Quadrat seiner Norm ergibt. Zur Erklärung des inneren Produktes wird darum einer der beiden Vektoren transponiert und seine Elemente werden komplex konjugiert $\mathbf{a}^H = [a_1^*, \dots, a_1^*]^T$ (oder „hermitisch“ transponiert). Das innere Produkt ergibt sich zu $\langle \mathbf{a}\mathbf{b} \rangle = \mathbf{a}^H \mathbf{b}$.

Damit sind die hermitische Matrix ($\mathbf{M}^H = \mathbf{M}$) und die unitäre Matrix ($\mathbf{M}^H \mathbf{M} = \mathbf{I}$) die komplexwertigen Analoga der symmetrischen und der orthogonalen Matrix. Zwei Vektoren sind unitär, wenn $\mathbf{a}^H \mathbf{b} = 0$.

Die betrachteten Quellensignale seien statistisch unabhängig. Das inkludiert, daß auch die statistischen Mittelwerte zweiter Ordnung verschwinden, also keine Kreuzkovarianzen auftreten. Die Kovarianzmatrix der Quellensignale, die durch die Mittelwertfreiheit mit der Korrelationsmatrix $\mathbf{R}_s^b = E\{\mathbf{s}^b \mathbf{s}^{bT}\}$ identisch ist, ist eine Diagonalmatrix.

Aus der Annahme, daß die Quellensignale statistisch unabhängig sind, folgt ohne weiteres, daß diese Eigenschaft auch für alle ihre schmalbandigen Komponenten gilt. Entsprechend kann man für jede Frequenz eine $M \times M$ Kovarianzmatrix bilden, die jeweils eine Diagonalmatrix ist und die Leistungen der einzelnen Signale im entsprechenden Frequenzband auf der Hauptdiagonalen enthält. Die Kovarianzmatrizen vereinigen also die Auto- und Kreuzleistungsdichtespektren der breitbandigen Schallsignale.

$$\mathbf{R}_s = E\{\mathbf{s}\mathbf{s}^H\} = \mathbf{\Lambda}_s \begin{bmatrix} E\{s_1 s_1^*\} & \dots & E\{s_1 s_M^*\} \\ \vdots & \ddots & \vdots \\ E\{s_M s_1^*\} & \dots & E\{s_M s_M^*\} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_M^2 \end{bmatrix} \quad (4.8)$$

Durch die Überlagerung der einzelnen Quellensignale sind die Mikrofonsignale hingegen miteinander korreliert. Deren $N \times N$ Kovarianzmatrix $R_x = E\{\mathbf{x}\mathbf{x}^H\}$ ist entsprechend ihrer Definition ebenfalls eine hermitische Matrix. Damit hat sie N reelle nicht-negative Eigenwerte $\lambda_1, \dots, \lambda_N$, die in der Summe der Gesamtleistung aller Observationen entsprechen. Weiterhin sind damit die zugehörigen Eigenvektoren $\mathbf{u}_1, \dots, \mathbf{u}_N$ unitär. Die Matrix aller Eigenvektoren sei $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_N]$ und die Diagonalmatrix mit den Eigenwerten auf der Hauptdiagonalen $\mathbf{\Lambda}_x = \text{diag}[\lambda_1, \dots, \lambda_N]$.

4.2 Konventionelle Beamformer

Konventionelle Beamformer lassen sich in zwei Ansätze unterteilen: Delay-&-Sum-Beamformer (DBF) und Filter-&-Sum-Beamformer (FBF). Im ersten Fall wird das Ausgangssignal aus der gewichteten Summe der zeitverzögerten Mikrofonsignale gebildet (Abb. 4.2a). Hier können nur die Gewichte w_n und die Zeitverzögerungen $\hat{\tau}_n$ modifiziert werden. Eine Erweiterung ist eine Überlagerung der gefilterten Signale (Abb. 4.2b). Dadurch können die Filterkoeffizienten w_{nk} einer beliebigen Länge K variiert werden.

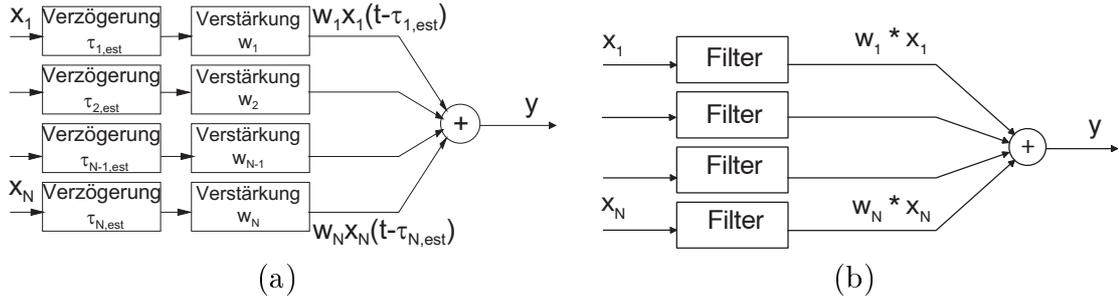


Abb. 4.2: Konventionelle Beamformer: Delay-&-Sum-Beamformer (a) und Filter-&-Sum-Beamformer (b)

4.2.1 Delay-&-Sum-Beamformer

Die Funktionsweise des Delay-&-Sum-Beamformers kann folgendermaßen formalisiert werden:

$$y^b(t) = \sum_{n=1}^N w_n^b x_n^b(t - \hat{\tau}_n). \quad (4.9)$$

$\hat{\tau}_n$ wird dabei so gewählt, daß die Verzögerungen ausgeglichen werden, die durch die unterschiedliche Laufzeit des zu rekonstruierenden Signals s_1 aus der Einfallrichtung ξ_1 zu den Mikrofonen entstehen (Fokussierung), siehe Abb. 4.1. Damit wird erreicht, daß diese Signale phasengleich aufaddiert werden und eine konstruktive Interferenz erfolgt. w_m^b ist die Shadingfunktion (z.B. eine typische Fensterfunktion), sie bestimmt die Unterdrückung der Signale aus anderen Richtungen.

Um das Verhalten des DBF zu erläutern und einige Begriffe einzuführen, wird die Wirkung des Beamformers auf das in (4.2) definierte Wellenfeld unter Freifeld- und Fernfeldbedingungen untersucht. Für monochromatische Wellen kann man die Wirkung des Beamformers leicht erkennen. Ein Signal s_m aus einer beliebigen Richtung $\vec{\xi}$ erreicht mit den Verzögerungen τ_{nm} die Mikrofone. Damit verändert sich (4.9) zu

$$\begin{aligned} y(t) &= \sum_{n=1}^N w_n^b s(t - \hat{\tau}_n + \tau_{nm}) \\ &= \sum_{n=1}^N w_n^b s\left(t - \frac{\vec{\xi}_1 \vec{x}_n}{c} + \frac{\vec{\xi}_m \vec{x}_n}{c}\right) \\ &= \sum_{n=1}^N w_n^b e\left(j\omega\left(t - \frac{\vec{\xi}_1 - \vec{\xi}_m}{c} \vec{x}_n\right)\right) \\ &= e^{j\omega t} \cdot \sum_{n=1}^N w_n^b e^{j\omega \frac{\vec{\xi}_1 - \vec{\xi}_m}{c} \vec{x}_n}. \end{aligned} \quad (4.10)$$

Das diskrete Arraypattern ist nach (3.20) die Fouriertransformation der Shadingfunktion: $W(\vec{k}) = \sum w_m^b$. Mit ihm ergibt sich

$$y(t) = e^{j\omega t} \cdot W\left(\omega \frac{\vec{\xi}_1 - \vec{\xi}_m}{c}\right).$$

Darin wird $W(\omega, \vec{\xi}_1, \vec{\xi}) = W(\frac{\omega \vec{\xi}_1}{c} - \vec{k})$ als Beampattern bezeichnet. Es beschreibt die Antwort des DBF, der auf die Richtung $\vec{\xi}_1$ eingestellt ist, auf Wellen aus Richtung $\vec{\xi}_m$. Ein typisches Arraypattern für ein Linienarray mit konstanter Gewichtung (Rechteckfunktion als Shadingfunktion) wurde bereits hergeleitet (vgl. (3.21)). In diesem Beispiel ist es eine si-Funktion (Abb. 3.3). Durch eine andere Shadingfunktion in Form eines von-Hann- oder Flat-Top-Fensters erhält man die zugehörigen spektralen Verläufe, die im folgenden als si-ähnliche Funktionen bezeichnet werden.

Das Beampattern ist das um $\vec{\xi}_1$ verschobene Arraypattern des Mikrofon-Arrays, so daß die Richtung der maximalen Durchlässigkeit durch diesen Parameter verschiebbar ist. Das Array wird also „gedreht“. Wellen aus Richtung ξ_1 werden ungehindert durchgelassen.

Die Signale aus anderen Richtungen werden entsprechend dem Arraypattern gedämpft. Da Arraypattern immer si-ähnliche Funktionen sind, existieren Nullstellen, d.h. Richtungen, in denen Signale maximal gedämpft werden. Die Dämpfung steigt, wenn sich die Richtung des einfallenden Signals von der Fokussierungsrichtung zunehmend unterscheidet. Werden die Anforderungen an die Mikrofonabstände nicht beachtet, tritt ein räumliches Aliasing auf, d.h. es bilden sich weitere Maxima im Beampattern aus.

Theoretisch ist es möglich, die Nullstellen des Beampatterns durch eine geeignete Wahl von w_n auf bestimmte Richtungen zu legen, und so Signale aus diesen Richtungen auszublenden. Da das Beampattern jedoch eine direkte Verknüpfung zwischen ξ und ω besitzt, verändert es sich mit jeder Frequenz, vgl. Abb. 4.3. Die Abbildung stellt die Dämpfung in Abhängigkeit von der Frequenz und der Richtung dar. Dunkle Gebiete bedeuten eine schwache und helle eine starke Dämpfung. Anhand der eingezeichneten Linie kann man die Wirkung des Beamformer auf Signale aus dieser Richtung erkennen: Sie werden nicht gleichmäßig gedämpft. Einen Nullbeamformer, der Signale einer Richtung unterdrückt, kann man also mit einem DBF nicht entwerfen. Versuche, die Frequenzabhängigkeit zu verringern, wurden von Sydow (1996) vorgeschlagen. Generell läßt sich ein Nullbeamformer jedoch nur mit einem FBF realisieren.

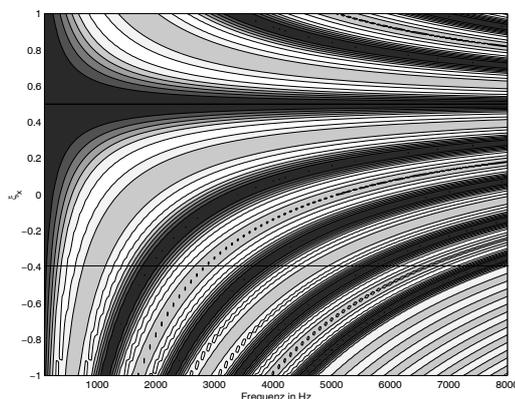


Abb. 4.3: Wellenzahl-Frequenz-Antwort eines DBF

Durch die Diskretisierung können die Verzögerungen $\hat{\tau}_n$ nur in diskreten Schritten

eingestellt werden. Damit ergibt sich eine Differenz zwischen der eingestellten $\vec{\xi}_1$ und der tatsächlichen „Sehrichtung“ $\vec{\xi}_1^*$ des Beamformers. Das Signal aus der Richtung $\vec{\xi}_1$ erfährt also eine Filterung nach $W(\omega \frac{\vec{\xi}_1 - \vec{\xi}_1^*}{c})$, d.h. es wird frequenzabhängig gefiltert (verzerrt) und gedämpft. Um das Problem zu beheben, schlagen Johnson und Dudgeon (1993) ein Upsampling der Signale vor, so daß die Verzögerungen genauer eingestellt werden können.

Das Problem kann man ebenfalls durch eine Verarbeitung im Frequenzbereich lösen. Wendet man die Definitionsgleichung des DBF auf die durch die STFT (4.3) transformierten Mikrosignale an, erhält man unter Ausnutzung der vektoriellen Notation:

$$\begin{aligned} y(t) &= \sum_{n=1}^N w_n^b x_n(t - \hat{\tau}_n) \\ &= \sum_{n=1}^N w_n^b \cdot x_n(t) \cdot e^{-\frac{\xi_n}{c} \hat{x}_n} \\ &= \hat{\mathbf{h}}_1^H \cdot \mathbf{W} \cdot \mathbf{x}^f(t) = \mathbf{w}^H \mathbf{x}^f(t). \end{aligned} \quad (4.11)$$

Offensichtlich entspricht die Wirkung des DBF einer Multiplikation mit dem in (4.6) definierten Verzögerungsvektor $\mathbf{h}_1(\vec{\xi}_0)$ und der Shading-Matrix mit den Gewichtungsfaktoren w_m der Mikrofone auf der Hauptachse.

$$\mathbf{W} = \begin{bmatrix} w_1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & w_M \end{bmatrix}$$

Oft wird $\mathbf{w}^H = \mathbf{h}_1^H \mathbf{W}$ als Steeringvektor bezeichnet.

Dieser Beamformer ist aufgrund seiner Einfachheit sehr robust, jedoch wegen seiner mangelnden Anpassungsfähigkeit fast nur noch in der Vorverarbeitung eingesetzt. So findet er z.B. für die Sprachverarbeitung als Vorverarbeitung für die weitere einkanalige Signalverarbeitung (Drews (1997), Bouquin-Jeannes *et al.* (1997)) oder als Vorverarbeitung für die ICA (Baumann *et al.* (2001)) Anwendung. Verbesserungen für Räume mit Nachhall wurden von Nishura *et al.* (2001) vorgeschlagen, die zusätzliche Spiegelquellen des Nutzsignals schätzen und diese in das Ergebnis integrieren. Wie in Johnson und Dudgeon (1993) gezeigt wird, ist die Wirkung eines DBF auf isotropen Schall eine Dämpfung von maximal \sqrt{N} . Dies stellt für inkohärenten Schall die Obergrenze für alle Beamformer dar.

4.2.2 Filter-&-Sum-Beamformer

Um einen Nullbeamformer (oder Constant Directivity Beamformer, vgl. Ward *et al.* (2001)) zu erreichen, muß man den Steeringvektor für jede Frequenz einzeln anpassen können. Das gelingt, indem man nicht nur zeitverzögerte sondern gefilterte Signale aufaddiert und die Filter angepaßte Frequenzgänge haben (Abb. 4.2). Solche

Beamformer auf der Basis von Transversalfiltern wurden unter anderem von Zelinski (1996) vorgeschlagen. Damit erweitert sich (4.9) zu

$$y^b(t) = \sum_{n=1}^N \sum_{k=0}^{K-1} w_{nk}^b x_n^b(t-k). \quad (4.12)$$

Wendet man auf (4.12) die STFT (4.3) an, ergibt sich

$$\begin{aligned} y^f(t) &= \sum_{n=1}^N w_n^f(f) x_n^f \\ &= \mathbf{w}^H(f) \cdot \mathbf{x}. \end{aligned} \quad (4.13)$$

Dabei ist $w_n^f(f)$ die äquidistante Abtastung der z-Transformation des in (4.12) definierten linearen Systems w_n^b auf dem Einheitskreis. Die üblichen Voraussetzungen der Kurzzeit-Fouriertransformation werden als erfüllt angenommen, insbesondere sei die Fensterlänge viel größer als die Filterlänge.

Überführt man (4.13) in die Vektorschreibweise, wird deutlich, daß \mathbf{w} der Steeringvektor ist. Er kann nun von der Frequenz unabhängig gewählt werden, d.h. man kann für jede zeitliche Frequenz ein eigenes Beampattern durch geeignete Filterkoeffizienten wählen. Bei einem so konstruierten Beamformer wird der feste Zusammenhang zwischen ω und $\vec{\xi}$ im Beampattern $W(\omega, \hat{\xi}_1, \hat{\xi})$ aufgehoben. Die Nullstellen können daher bei jeder Frequenz frei gewählt und damit so gelegt werden, daß sie bei bestimmten Richtungen $\vec{\xi}_k$ liegen. Diese Beamformer können Signale aus bestimmten Richtungen ausblenden („zu Null setzen“) und heißen folglich Nullbeamformer. Sie besitzen gegenüber DBF's für kohärenten Schall eine große Überlegenheit. Sind die Richtungen der Nutz- und Störquellen bekannt und übersteigt die Anzahl der Quellen nicht die Anzahl der möglichen Nullstellen des Arraypattern, stellt ein solcher Nullbeamformer ein Optimum dar, das mit einem adaptiven Beamformer oder der blinden Quellentrennung im besten Fall (optimal geschätzte Signalstatistik) erreicht, aber nicht übertroffen werden kann (Kajala und Hämäläinen (2001), Araki *et al.* (2001)).

Die Gleichung (4.13) kann man auch direkt implementieren, indem man das Signal in schmalbandige Komponenten zerlegt, (4.13) anwendet und das breitbandige Signal rekonstruiert. Ein Beispiel für dieses Vorgehen ist in Elko (1996) genau angegeben. Es wird auch in den folgenden Kapiteln für adaptive Beamformer und die blinde Quellentrennung eingesetzt. Ähnlich sind Subband-Verfahren. Bei ihnen wird das breitbandige Signal in Komponenten transformiert, die nicht so schmalbandig sind, daß sie als Aufnahmen monochromer Wellen aufgefaßt werden können. Auf jedes Frequenzband wird ein separater DBF angewendet (z.B. McCowan und Sridharan (2001)), wodurch sich ebenfalls ein frequenzabhängiges Arraypattern realisieren läßt. Weitere aktuelle Beispiele für FBF-Realisierungen sind Kajala und Hämäläinen (2001), Kurpjuhn *et al.* (2001) und Wang *et al.* (2001).

Betrachtet man beide Beamformer im Frequenzbereich, (4.11) und (4.13), gehen sie ineinander über. Der einzige Unterschied besteht darin, daß beim DBF N Verzögerungen und ebenso viele Gewichte für die Shadingfunktion einzustellen sind,

während der FBF $N \cdot K$ Filterkoeffizienten als Parameter hat. In der großen Anzahl liegt auch der Nachteil eines FBF gegenüber einem DBF.

Eine Vereinfachung kann man erreichen, wenn man die gleichen Filterkoeffizienten auf alle Sensoren anwendet: $w_{nk} = w_k$. Das entspricht einem DBF mit nachgelagerter einkanaliger Filterung (Drews (1999), Bouquin-Jeannes *et al.* (1997)). Aktuelle Implementierungen stellen bereits einen Übergang zu adaptiven Beamformern dar, weil zur Bestimmung der Filterkoeffizienten die Informationen aus den unterschiedlichen Kanälen, insbesondere die Auto- und Kreuzleistungsdichten der breitbandigen Signale, genutzt werden. Dazu werden meist Wiener-Filter (siehe Haykin (1996)) oder Modifikationen von Wiener-Filtern (Ephraim-Malah-Filter) eingesetzt. Da die Leistungsdichten der breitbandigen Signale den Kovarianzen der schmalbandigen Komponenten entsprechen, sind diese Algorithmen eng mit den in Kap. 4.3 besprochenen verbunden. Der Unterschied besteht in einer geringeren Anzahl einstellbarer (und damit zu schätzender) Parameter.

4.3 Adaptive Beamformer

Adaptive Beamformer (ABF) passen sich den aufgenommenen Signalen an, d.h. sie verändern ihre Richtcharakteristik in Abhängigkeit von den Signalen. Meistens werden sie als FBF mit veränderlichen Filterkoeffizienten realisiert. Damit lassen sich das Beampattern und die Nullstellen sowie die Hauptdurchlaßrichtung einstellen.

Von den zahlreichen in der Literatur vorgeschlagenen adaptiven Beamformern basieren die meisten zur Rekonstruktion des Signals eingesetzten Beamformern auf einer Minimum-Variance- oder Maximum-Likelihood-Schätzung. Nach ihren Urhebern oder ihren Grundprinzipien haben sich unterschiedliche Namen etabliert. Für einen Beamformer mit einer Minimum-Variance-Schätzung sind die häufigsten Namen „Frost-Beamformer“ oder „Minimum Variance Beamformer“⁴, wenn nur die Einfallsrichtung des Nutzsignals als Vorwissen bekannt sein muß (Frost (1972)). Benötigt man zusätzlich noch die „Nullrichtungen“, aus denen die Störsignale kommen (Griffith und Jim (1982)), existieren weiterhin die Namen „Linearly constrained minimum variance beamformer“, „Griffith-Jim-Beamformer“ oder „Generalized sidelobe beamformer“. Beamformer auf Basis einer Maximum-Likelihood-Schätzung haben die Namen „Capon's method“ oder „Maximum Likelihood Beamformer“ (Capon (1967)). Häufig ist in der Literatur die Namensgebung nicht konsistent, so daß man nicht immer ohne weiteres erkennen kann, um welches Grundprinzip es sich bei einem konkreten Algorithmus handelt.

Adaptive Beamformer reagieren sehr sensibel auf Parameterschwankungen. Wie von Affes und Grenier (1994) empirisch und von Bitzer *et al.* (1999) theoretisch gezeigt, können erhebliche Verzerrungen auftreten, wenn die Anwendungsbedingungen suboptimal sind. Die wesentlichen Gründe dafür sind:

⁴Dieser Ausdruck wird u.a. von Johnson und Dudgeon (1993) kritisiert, weil die Varianz nur für stochastische Signale definiert ist. Von ihnen wird deshalb die Bezeichnung „Minimum Power Beamformer“ vorgeschlagen. Da sich der Begriff Minimum Variance dennoch durchgesetzt hat, soll er auch in dieser Arbeit benutzt werden.

1. Nachhall, der durch eine Ausbreitung der Schallsignale auf mehreren Wegen entsteht (Mehrwegeüberlagerung),
2. Lokalisationsfehler des Nutzsignals, d.h. eine schlecht geschätzte Einfallsrichtung, und
3. Probleme der digitalen Signalverarbeitung (begrenzte Wortlänge, begrenzte Auflösbarkeit von Zeitverzögerungen).

Bis vor kurzem wurde darum vom Einsatz adaptiver Beamformer in der Praxis gänzlich abgeraten (z.B. Drews (1999)).

In diesem Kapitel soll das Verhalten der adaptiven Beamformer quantitativ und analytisch untersucht werden, wenn die „Problemfälle“ eintreten. Sie sind häufige Fehlerursache, Nachhall ist in geschlossenen Räumen unvermeidlich, und in vielen Fällen kann die Einfallsrichtung nicht exakt angegeben werden. Die Quantifizierung der Performance-Einbrüche wurde in den o.g. Arbeiten nur angerissen, so daß sie mit diesem Formalismus noch nicht existiert. Für die vorliegende Arbeit ist sie jedoch von gewissem Interesse, weil sie die Frage beantwortet, ob adaptive Beamformer in der akustischen Güteprüfung eingesetzt werden können. Außerdem wird der entwickelte Formalismus für die Bewertung der neuen Algorithmen im weiteren Verlauf benutzt.

4.3.1 Minimum Variance Beamformer

Der Minimum Variance (MV) Beamformer geht auf Frost (1972) zurück und wurde seit seiner Einführung weiterentwickelt, ohne daß sich seine ursprüngliche Wirkungsweise dadurch verändert hat. Die größte Erweiterung wurde von Griffith und Jim (1982) eingeführt: Sie schlugen vor, explizit Nullen des Arraypatterns auf bestimmte Richtungen zu legen („Generalized sidelobe canceller“). Zusammengefaßt sind die Konzepte auch unter dem Begriff „Linearly constrained minimum variance“ Beamformer bekannt. Für eine Übersicht über weitere Beiträge siehe Johnson und Dudgeon (1993), Haykin (1996) oder Bitzer und Simmer (2001).

Die Basis ist ein FBF im Frequenzbereich (vgl. Kap. 4.2.2), dessen Steeringvektor $\mathbf{w}(f)$ adaptiert wird. Die Grundidee ist eine Minimierung der Ausgangsleistung (Ausgangsvarianz) des Beamformers

$$J = E\{y^2\} = E\{\mathbf{w}^T \mathbf{xx} \mathbf{w}^T\} = \mathbf{w}^T E\{\mathbf{xx}\} \mathbf{w}^T = \mathbf{w}^T \mathbf{R}_x \mathbf{w} \quad (4.14)$$

mit der Randbedingung, daß Signale, die mit dem Verzögerungsvektor des Nutzsignals multipliziert wurden, ungehindert durchgelassen werden.

$$\mathbf{w}^H \hat{\mathbf{h}}_1 = c_1 \quad (4.15)$$

Diese Bedingung, die zwingend Vorwissen über die Signale und ihre Mischung erfordert, ist notwendig, denn das globale (unbedingte) Minimum von (4.14) ist der Koordinatenursprung ($\mathbf{w} = \mathbf{0}$).

Die bedingte Optimierung erfolgt wie üblich über Lagrange- Multiplikatoren (Bronstein *et al.* (2000)), die zu optimierende Funktion (Lagrangefunktion) ist also

$$J_{ABF} = E[y^2] + \alpha(\mathbf{w}^H \hat{\mathbf{h}}_1 - c_1).$$

Die Optimierung soll im folgenden mit Streudiagrammen in Abb. 4.5 und 4.4 verdeutlicht werden. Streudiagramme sind zweidimensionale Häufigkeiten einer Stichprobe durch die Dichte einer Punktwolke dargestellt, die bei genügend großer Stichprobe die zweidimensionale Wahrscheinlichkeitsdichtefunktion repräsentieren. Daher bedeutet eine hohe Punktdichte eine hohe Auftretenswahrscheinlichkeit.⁵

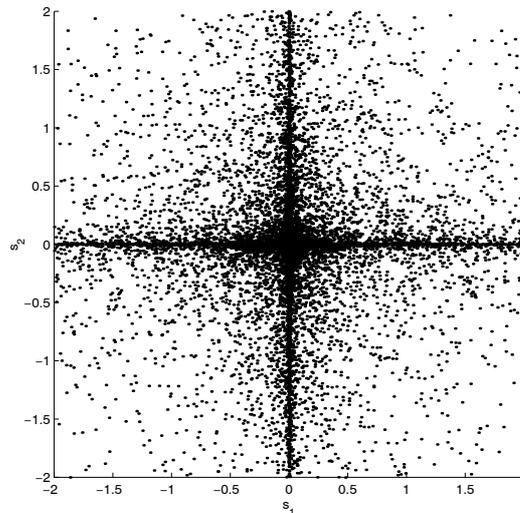


Abb. 4.4: Streudiagramm der Quellensignale

Man erkennt in Abb. 4.4, daß keine statistische Abhängigkeit zwischen den Quellensignalen herrscht. In Abb. 4.5 sieht man hingegen die starke Korrelation zwischen den Observationssignalen x_1 und x_2 . Weiterhin kann man zwei Bereiche starker Häufung unterscheiden. Legt man Ausgleichsgeraden durch die einzelnen Punktwolken, haben diese die Richtungen der Mischungsvektoren \mathbf{h}_1 und \mathbf{h}_2 .

Weiterhin sind in dieser Abbildung die Höhenlinien der in (4.14) definierten Energie J eingezeichnet. (4.14) beschreibt deshalb eine Energie, weil die quadratische Form einer positiv semi-definiten Matrix nicht-negativ ist und jede Kovarianzmatrix hermitisch und positiv semi-definit ist. Die Energiefunktion hat die Form einer Schüssel („bowl shape“): Ihr Minimum liegt bei $\mathbf{0}$, und sie wächst mit dem Betrag von \mathbf{w} . Für einen konstanten Betrag erreicht sie ihr Minimum und Maximum in den Richtungen der Eigenvektoren von \mathbf{R}_x , welche in Richtung der größten Korrelation zeigen bzw. senkrecht darauf stehen.

⁵Obwohl die M - bzw. N -dimensionalen Zufallsvektoren \mathbf{s} und \mathbf{x} komplexwertig sind, sind aus zeichnerischen Gründen in dieser Arbeit nur Streudiagramme mit zweidimensionalen reellwertigen Vektoren dargestellt. Die Aussagen gelten jedoch auch für höherdimensionale Räume analog.

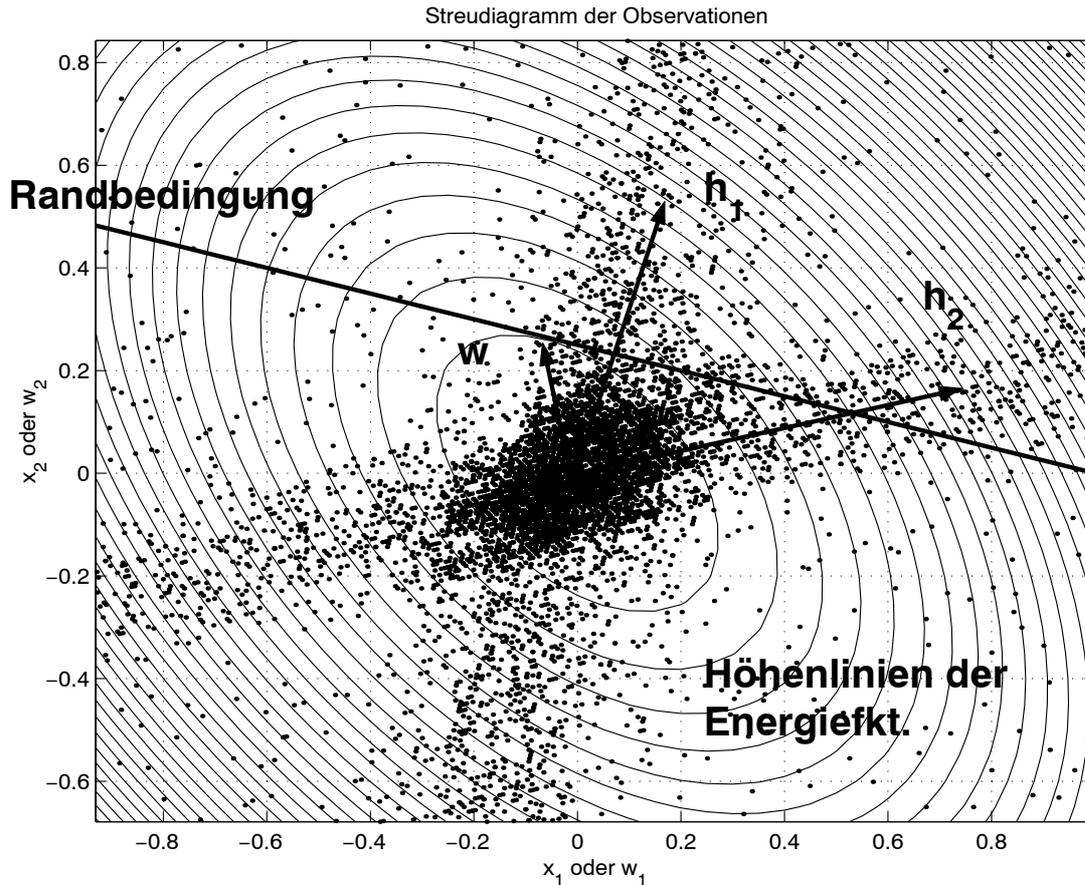


Abb. 4.5: Streudiagramm der aufgenommenen Signale mit den Höhenlinien der Energiefunktion und der Randbedingung.

Formt man die Randbedingung (4.15) in die Hessesche Normalform um, kann man sie ebenfalls als Gerade (oder im generellen Fall als $(N - 1)$ -dimensionale Hyperebene) in die Graphik einzeichnen.

Die Optimierung erfolgt nun auf der Geraden der Randbedingung. Man sucht nach dem Punkt, an dem sie die „engste“ Höhenlinie berührt. Formal erfolgt die Ableitung mit Hilfe der Kuhn-Tucker-Bedingungen, siehe Bronstein *et al.* (2000). Diese haben die Voraussetzung, daß eine Funktion $J(\mathbf{x})$ unter den Nebenbedingungen $g_i(\mathbf{x}) \leq 0$ ($i \in \mathbf{I}$) minimiert werden soll. Dann genügt ein Vektor \mathbf{x}_0 den Kuhn-Tucker-Bedingungen g.d.w.

$$\nabla_{\mathbf{x}} J(\mathbf{x}_0) = \sum_{i \in \mathbf{I}} \alpha_i \nabla_{\mathbf{x}} g_i(\mathbf{x}_0) \quad \alpha_i \geq 0.$$

Ein Kuhn-Tucker-Vektor ist ein stationärer Punkt der Lagrangefunktion und damit eine potentielle Lösung des bedingten Optimierungsproblems. Reduziert sich die Anzahl der Randbedingungen auf eine, vereinfachen sich die Kuhn-Tucker-Bedingungen zu

$$\nabla_{\mathbf{x}} J(\mathbf{x}_0) = \alpha \nabla_{\mathbf{x}} g(\mathbf{x}_0).$$

Geometrisch betrachtet heißt das für einen stationären Punkt, daß der Gradient der Energiefunktion J senkrecht auf der Nebenbedingung stehen muß. Dann hat der Gradient keine Richtungskomponente in der Hyperebene der Randbedingung, so daß eine Bewegung bei Einhaltung der Nebenbedingung zu keiner Verringerung der Energie führt.

Im konkreten Optimierungsproblem ist $J(\mathbf{w})$ eine reell-wertige skalare Funktion des komplexen Vektors \mathbf{w} . Solche Funktionen erfüllen nicht die Cauchy-Riemannschen Bedingungen (Bronstein *et al.* (2000)) und sind damit nicht analytisch, so daß die Ableitungen nicht definiert sind. Die in der Signalverarbeitung übliche Abhilfe ist der komplexe Gradient, der von Brandwood (1983) vorgeschlagen wurde (siehe auch Therrien (1992), Viberg (2002), u.a.).

Dazu wird $J(z, z^*)$ als Funktion zweier unabhängiger Veränderlicher betrachtet, so daß sich komplexe Gradienten $\nabla_z J = \left(\frac{\partial}{\Re(z)} + j \frac{\partial}{\Im(z)} \right) J$ und $\nabla_{z^*} J = \left(\frac{\partial}{\Re(z)} - j \frac{\partial}{\Im(z)} \right) J$ definieren lassen. Wenn das Argument skalar ist, ist für einen stationären Punkt von $J(z, z^*)$ das Verschwinden einer der beiden Gradienten notwendig und hinreichend.⁶ Ist dazu das Argument ein Vektor \mathbf{z} , muß der konjugiert-komplexe Gradient $\nabla_{\mathbf{z}^*} = [\nabla_{z_1^*}, \dots, \nabla_{z_T^*}]^T$ betrachtet werden⁷.

Der Gradient der Energiefunktion ist $\nabla_{\mathbf{w}^*} J = \nabla_{\mathbf{w}^*} \mathbf{w}^H \mathbf{R}_x \mathbf{w} = \mathbf{R}_x \mathbf{w}$, der Gradient der Hyperebene $\nabla_{\mathbf{w}^*} (\mathbf{w}^H \mathbf{h}_1 - c_1) = \mathbf{h}_1$. Damit ergibt sich für den Kuhn-Tucker-Punkt

$$\mathbf{R}_x \mathbf{w} = \alpha \mathbf{h}_1 \quad \Rightarrow \quad \mathbf{w} = \alpha \mathbf{R}_x^{-1} \mathbf{h}_1.$$

Setzt man diese Bedingung in die Gleichung der Nebenbedingung (4.15) ein, ergibt sich für $\alpha = c_1 \mathbf{h}_1^H \mathbf{R}_x^{-1} \mathbf{h}_1$, womit man den optimalen Vektor zu

$$\mathbf{w}_{\text{MV}} = \frac{c_1 \mathbf{R}_x^{-1} \hat{\mathbf{h}}_1}{\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \hat{\mathbf{h}}_1}. \quad (4.16)$$

erhält. Im folgenden wird für die vereinfachte Notation die Konstante $c_1 = 1$ gesetzt.

Wie in Abb. 4.5 zu sehen (und in Kap. 4.4.3 auch formal gezeigt), steht der so gewonnene Vektor orthogonal auf den Störsignalen ($p_1 = \mathbf{w}_{\text{MV}}^H \hat{\mathbf{h}}_2 = 0$). Damit ist er auch optimal im Sinne eines linearen MMSE-Schätzers, der die Orthogonalität des geschätzten Signals zu allen Linearkombinationen der Störungen als Optimierungskriterium besitzt.

4.3.2 Maximum Likelihood Beamformer

Der Maximum Likelihood (ML) Beamformer wurde von Capon (1967) eingeführt und basiert auf einer Maximum Likelihood Schätzung des Signals s_1 für eine be-

⁶Das kann man leicht sehen, wenn man das erste Glied der Taylor-Reihe für eine kleine Änderung Δz betrachtet: $\Delta J = (\nabla_z J) \Delta z + (\nabla_{z^*} J) \Delta z^* = 2\Re\{\nabla_z J \Delta z\} = 2\Re\{\nabla_{z^*} J \Delta z^*\}$. Soll sich J für z_0 nicht ändern, reicht es zu zeigen, daß $\nabla_z J = 0$ oder $\nabla_{z^*} J = 0$.

⁷Das wird ebenfalls über die Taylorreihe (für $\Delta \mathbf{z}$) gezeigt: $\Delta J = \sum_i ((\nabla_{z_i} J) \Delta z_i + (\nabla_{z_i^*} J) \Delta z_i^*) = \sum_i 2\Re\{(\nabla_{z_i} J) \Delta z_i\} = 2\Re\{(\nabla_{\mathbf{z}} J)^T \Delta \mathbf{z}\} = 2\Re\{(\nabla_{\mathbf{z}^*} J)^H \Delta \mathbf{z}\}$. Die Richtung der stärksten Änderung zeigt also nach \mathbf{z}^* .

kannte Verteilung des Störung. Dazu wird das Signal gesucht, bei dem der Observationsvektor am wahrscheinlichsten auftritt.

$$\max_{s_1} \ln(p_{\mathbf{x}|s_1})$$

Die erste Annahme dieses Beamformers ist, daß die Störsignale in ihrer Summe normalverteilt sind. Dadurch lassen sie sich durch folgende Wahrscheinlichkeitsdichtefunktion beschreiben. $\tilde{\mathbf{x}}$ ist darin der Observationsvektor, wenn kein Nutzsignal enthalten ist. Ebenso ist $\mathbf{K}_x = E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\}$ die Kovarianzmatrix der Störsignale und des Rauschens:

$$p_{\tilde{\mathbf{x}}} = \frac{1}{\sqrt{2\pi|\det \mathbf{K}_x|}} e^{-\frac{1}{2}\tilde{\mathbf{x}}^H \mathbf{K}_x^{-1} \tilde{\mathbf{x}}}. \quad (4.17)$$

Der Anteil der Störsignale am Observationsvektor kann durch die folgende Gleichung bestimmt werden:

$$\begin{aligned} \mathbf{x} = \sum_m \mathbf{h}_m s_m &= \sum_{m>1} \mathbf{h}_m s_m + \mathbf{h}_1 s_1 = \tilde{\mathbf{x}} + \mathbf{h}_1 s_1 \\ \tilde{\mathbf{x}} &= \mathbf{x} - \mathbf{h}_1 s_1. \end{aligned}$$

Wenn der Verzögerungsvektor wie in den vorhergehenden Fällen mit $\hat{\mathbf{h}}_1$ vorgegeben ist, kann man die Wahrscheinlichkeitsdichte und die logarithmische Likelihood-Funktion aufstellen:

$$\begin{aligned} p_{\mathbf{x}|s_1} &= \frac{1}{\sqrt{2\pi|\det \mathbf{K}_x|}} e^{-\frac{1}{2}(\mathbf{x} - \hat{\mathbf{h}}_1 s_1)^H \mathbf{K}_x^{-1} (\mathbf{x} - \hat{\mathbf{h}}_1 s_1)} \\ \ln p_{\mathbf{x}|s_1} &= \frac{1}{2}(\mathbf{x} - \hat{\mathbf{h}}_1 s_1)^H \mathbf{K}_x^{-1} (\mathbf{x} - \hat{\mathbf{h}}_1 s_1) - \frac{1}{2} \ln 2\pi|\det \mathbf{K}_x|. \end{aligned}$$

Für die Suche nach dem Maximum muß der komplexe Gradient nach s_1 gebildet und zu Null gesetzt werden:

$$\begin{aligned} 0 = \nabla_{s_1} \ln p_{\mathbf{x}|s_1} &= -\frac{1}{2} \hat{\mathbf{h}}_1^H \mathbf{K}_x^{-1} (\mathbf{x} - \hat{\mathbf{h}}_1 s_1) \\ \Leftrightarrow 0 &= \hat{\mathbf{h}}_1^H \mathbf{K}_x^{-1} \mathbf{x} - \hat{\mathbf{h}}_1^H \mathbf{K}_x^{-1} \hat{\mathbf{h}}_1 s_1, \end{aligned}$$

woraus sich leicht das gesuchte Signal \hat{s}_1 und damit der Steeringvektor schätzen lassen:

$$y = s_1 = \frac{\hat{\mathbf{h}}_1^H \mathbf{K}_x^{-1} \mathbf{x}}{\hat{\mathbf{h}}_1^H \mathbf{K}_x^{-1} \hat{\mathbf{h}}_1} \Rightarrow \mathbf{w}_{\text{ML}} = \frac{\mathbf{K}_x^{-1} \hat{\mathbf{h}}_1}{\hat{\mathbf{h}}_1^H \mathbf{K}_x^{-1} \hat{\mathbf{h}}_1}. \quad (4.18)$$

Der Steeringvektor des ML-Beamformer ist dem Vektor des MV-Beamformers sehr ähnlich. Der wesentliche Unterschied besteht darin, daß für den ML-Beamformer die Kovarianzmatrix der Störgeräusche (ohne Nutzsignale) bekannt sein muß.

Die Kovarianzmatrix der Nebengeräusche kann nur in den Pausen geschätzt werden, wenn das Nutzsignal nicht vorhanden ist. Das bringt zwei Einschränkungen mit sich:

1. Die Pausen müssen vorhanden sein und detektiert werden. Bei Sprachsignalen setzt das eine Aktivitätssteuerung voraus, die zusätzlichen Aufwand erfordert. Bei Maschinensignalen ist die Detektion dann unproblematisch, wenn man die Maschinen abschalten kann. Dann sind die Signalpausen klar definiert. In anderen Fällen treten keine Signalpausen auf.
2. Die Umgebungsgeräusche müssen über einen längeren Zeitraum stationär sein, weil \mathbf{K}_x zu unterschiedlichen Zeitpunkt geschätzt und verwendet wird. Diese Bedingung wird regelmäßig nicht erfüllt, denn abgesehen vom isotropen Rauschen sind alle anderen Störanteile sehr zeitabhängig und somit nicht stationär. Dieser Nachteil ist besonders bei der Analyse von Maschinensignalen unvorteilhaft, weil keine kurzzeitigen Pausen wie bei menschlicher Sprache auftreten. Darum liegt ein langer zeitlicher Abstand zwischen der Bestimmung von \mathbf{K}_x und der eigentlichen Messung (Einschaltvorgang der Apparatur, Hochlauf etc.), der länger als typische Stationaritätsannahmen (z.B. 20 ms bei Sprache) ist. Dies macht den Einsatz dieses Beamformers für die Maschinendiagnose unpraktikabel.

Trotz dieser Nachteile wird der ML-Beamformer häufig eingesetzt. Das häufigste Anwendungsgebiet der akustischen Signalverarbeitung ist die Sprachverarbeitung, bei der man die Kovarianzmatrix sehr zeitnah in den Sprachpausen schätzen kann. Das heißt in der Praxis, daß die Adaption meistens gestoppt wird, sobald das Nutzsignal vorhanden ist. Der Grund für seinen häufigen Einsatz ist der Vorteil, daß er deutlich weniger empfindlich gegenüber Parameterschwankungen ist, wie im folgenden Abschnitt gezeigt wird.

4.4 Einfluß einer ungenauen Schätzung des Verzögerungsvektors

4.4.1 Gründe für die ungenaue Schätzung

Der hauptsächliche Nachteil der zuvor beschriebenen adaptiven Beamformer besteht in ihrer extremen Abhängigkeit von der Schätzqualität des Verzögerungsvektors \mathbf{h}_1 . Im allgemeinen ist die Raumimpulsantwort nicht bekannt, so daß der Verzögerungsvektor⁸ unter Freifeldannahme geschätzt wird (4.6).

Zunächst müssen die Positionen der Mikrofone und der Nutzsignalquelle bekannt sein. Während die Mikrofonpositionen im allgemeinen hinreichend genau bestimmt werden können,⁹ ist dies für den Ort der Quelle schwieriger. Beispiele für die Be-

⁸Der Name „Verzögerungsvektor“ ist an dieser Stelle nicht mehr präzise. $\mathbf{h}_1(f)$ besteht aus den Fouriertransformierten der Raumimpulsantworten von der Nutzquelle zu den Mikrofonen, die bei Aufgabe der Freifeldannahme keine reinen Verzögerungen mehr sind. Dennoch ist der Name üblich und soll folgenden weiter gebraucht werden.

⁹Vgl. Johnson und Dudgeon (1993) oder Jabloun und Champagne (2001) für die Behandlung von Fehler, die durch ungenaue Mikrofonpositionen entstehen

stimmung werden in Kap. 4.1 angegeben. Dadurch können Fehler entstehen, deren Einfluß unter dem Stichwort „DOA mismatch“ behandelt wird.

Der zweite Grund für Schätzfehler ist der Nachhall, da (4.6) nur den Direktschall auswertet, während der Schall sich durch die Reflexionen auf unterschiedlichen Pfaden ausbreitet und damit mit unterschiedlichen Verzögerungen auf die Mikrofone trifft („Mehrwege-Mischung“). Das heißt, es können erhebliche Schätzfehler auftreten, obwohl die Quellen- und Mikrofonpositionen genau bestimmt sind.

Um diese Fehler genauer zu quantifizieren, wird der Fehlerwinkel des Verzögerungsvektors (Schätzfehlerwinkel, steering vector error angle SVA, μ) eingeführt. Dazu benutzen wir eine generalisierte Cosinus-Definition für komplexe Vektoren. Mit dem inneren Produkt $\langle \mathbf{x}, \mathbf{y} \rangle$ zweier komplexwertiger Vektoren $\mathbf{x}, \mathbf{y} \in \mathbb{C}^2$ wird der Cosinus als

$$\cos(\mathbf{x}, \mathbf{y}) = \frac{|\langle \mathbf{x}, \mathbf{y} \rangle|}{\sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \sqrt{\langle \mathbf{y}, \mathbf{y} \rangle}} \quad (4.19)$$

definiert. Der Schätzfehlerwinkel $\mu(f)$ sei nun der Winkel zwischen dem tatsächlichen Verzögerungsvektor einer Frequenz $\mathbf{h}_1(f)$ und seiner Schätzung $\hat{\mathbf{h}}_1(f)$ und $\alpha(f)$ der Winkel zwischen den Verzögerungsvektoren zweier Quellen (z.B. $\mathbf{h}_1(f)$ und $\mathbf{h}_2(f)$):

$$\mu(f) = \arccos \left(\frac{|\mathbf{h}_1^H \hat{\mathbf{h}}_1|}{|\mathbf{h}_1| |\hat{\mathbf{h}}_1|} \right) \quad \alpha(f) = \arccos \left(\frac{|\mathbf{h}_1^H \mathbf{h}_2|}{|\mathbf{h}_1| |\mathbf{h}_2|} \right). \quad (4.20)$$

An dieser Stelle soll betont werden, daß diese Winkel in der Streuebene (bzw. im Streuraum) der Vektoren definiert sind und etwas über deren statistische Eigenschaften aussagen. Sie sind frequenzabhängig und *nicht* identisch mit Winkeln im (physikalischen) Raum. Im Streudiagramm können die Winkel leicht definiert werden (vgl. Abb. 4.6a/b).

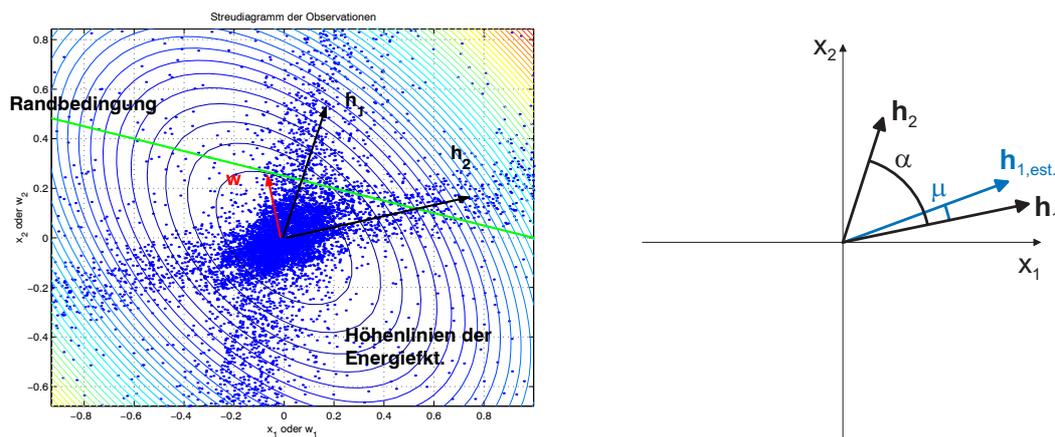


Abb. 4.6: Definition von α und μ im Streudiagramm

Um sich eine Vorstellung von der Größe dieser Winkel zu machen, wird folgende Simulation angenommen: Ein Mikrofon-Array aus zwei Mikrofonen in einem Abstand von 4 cm befindet sich in einem Freifeld (kein Nachhall) und ein Signal trifft in einem Einfallswinkel von $\theta = 30^\circ$ auf das Array. Die Abtastrate beträgt

$f = 8\text{kHz}$, und die Daten werden mit einer FFT-Länge von 1024 in den Frequenzbereich transformiert. Die Verzögerungsvektoren werden mit falschen Einfallsrichtungen geschätzt.

Das Ergebnis dieser Simulation ist für unterschiedliche Fehler in der Schrichtung des Arrays in Abb. 4.7 dargestellt. Der Fehlerwinkel μ steigt mit zunehmender Frequenz erwartungsgemäß an, weil sich eine falsch geschätzte Richtung in einer falschen Verzögerung $\hat{\tau}$ niederschlägt. Es ergibt sich also ein Fehler zwischen der realen und der geschätzten Verzögerung $\Delta\tau$ und damit

$$\begin{aligned} \mathbf{h}_1^H \hat{\mathbf{h}}_1 &= e^{-2\pi f \tau_{11}} e^{2\pi f \hat{\tau}_{11}} + e^{-2\pi f \tau_{12}} e^{2\pi f \hat{\tau}_{12}} = e^{2\pi f (\hat{\tau}_{11} - \tau_{11})} + e^{2\pi f (\hat{\tau}_{12} - \tau_{12})} \\ &= e^{2\pi f \Delta\tau_{11}} + e^{2\pi f \Delta\tau_{12}}, \end{aligned}$$

so daß sich der Fehlerwinkel proportional zur Frequenz erhöht. Der maximale Wert für μ in dieser Simulation liegt auch dann unter 20° , wenn die Einfallsrichtung sehr grob geschätzt wurde.

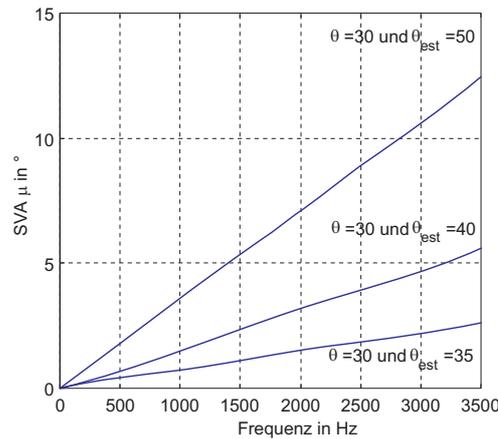
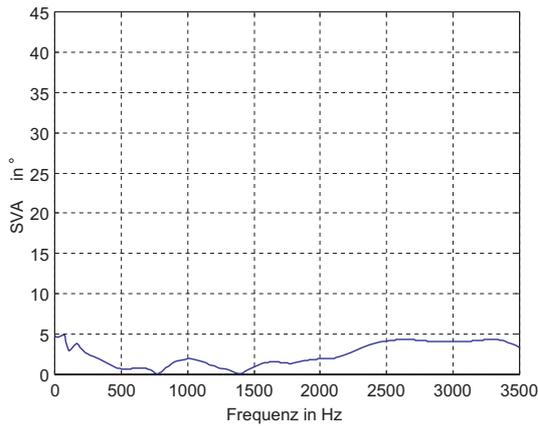


Abb. 4.7: Fehlerwinkel μ (Schätzfehlerwinkel) für die Simulation einer falsch geschätzten Einfallsrichtung

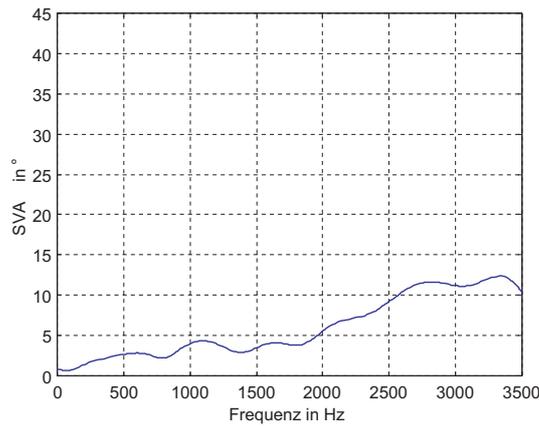
Das größere Problem wird durch Nachhall verursacht. Abb. 4.8 zeigt den Schätzfehlerwinkel für reale Räume¹⁰ mit Nachhallzeiten von $t_r = 150\text{ ms}$ und $t_r \leq 15\text{ ms}$ für die gleichen Parameter (FFT-Länge, Position der Mikrofone etc) wie oben.

Die Fehler durch Nachhall können also deutlich höhere Werte annehmen, als die durch einen „DOA-mismatch“ induzierten. Das ist besonders problematisch, weil Nachhall wird praktisch immer auftritt. Der Einfluß der Nachhalls übersteigt den der falschen Richtungsschätzung deutlich, so daß bei Nachhall nie mit einem korrekten Verzögerungsvektor gerechnet werden darf. Das folgende Kapitel zeigt den Einfluß eines unpräzisen Verzögerungsvektors.

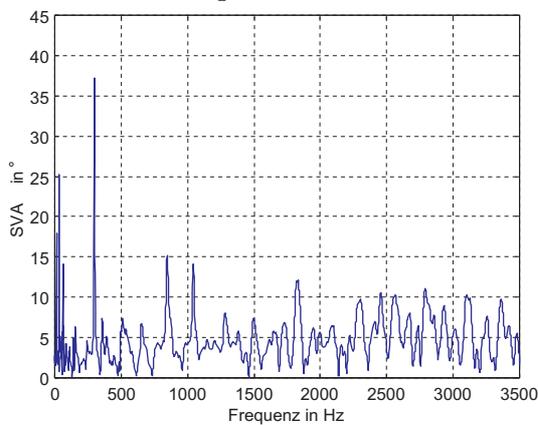
¹⁰Detailliert sind diese Räume in Kap. 8.1 beschrieben.



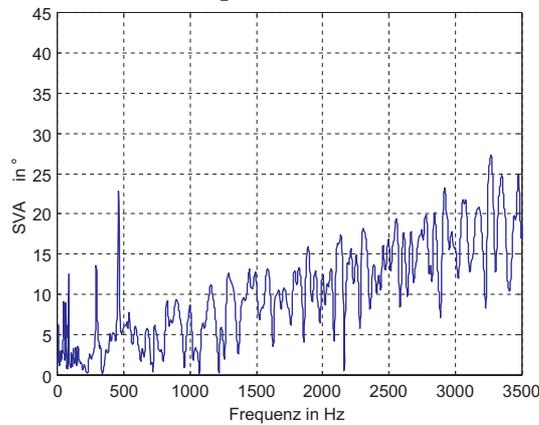
(a) Korrekte Schätzung der Einfallrichtung $\Delta\theta = 10^\circ$, $t_r \leq 15\text{ ms}$



(b) Inkorrekte Schätzung der Einfallrichtung $\Delta\theta = 10^\circ$, $t_r \leq 15\text{ ms}$



(c) Korrekte Schätzung der Einfallrichtung $\Delta\theta = 10^\circ$, $t_r = 150\text{ ms}$



(d) Inkorrekte Schätzung der Einfallrichtung $\Delta\theta = 10^\circ$, $t_r = 150\text{ ms}$

Abb. 4.8: Vergleich der durch grobe Schätzung der Einfallrichtung und Nachhall entstandenen Schätzfehlerwinkel.

4.4.2 Konsequenzen für den ML-Beamformer

Um die Robustheit der behandelten Beamformer analytisch zu bewerten, müssen einige vereinfachende Annahmen gemacht werden:

1. Die Anzahl der Mikrofone und Quellen ist $N = M = 2$.
2. Die beiden Quellen tragen die gleiche Leistung zum gemessenen Signal bei. Das ist gewährleistet, wenn die Quellen etwa gleich laut und ähnlich weit vom Array entfernt sind. Dann kann man festlegen, daß die Varianzen der Signale $\sigma_{s_1} = \sigma_{s_2} = \sigma_s$ und die Normen der Verzögerungsvektoren $|\mathbf{h}_1| = |\mathbf{h}_2| = 1$ sind. Die gesamte Energie liegt in den Quellensignalen, eine Abschwächung durch den Transport im Raum wird nicht berücksichtigt.

Die erste Annahme kann leicht „aufgeweicht“ werden, indem die Untersuchung für jeweils zwei Kanäle gemacht wird. Die zweite sagt, daß nur die Energie am Ort der Mikrofone geschätzt werden kann und soll. Eine Entfaltung findet nicht statt.

Das Ausgangssignal y des ML-Beamformers ergibt nach (4.18)

$$\begin{aligned}
 y &= \mathbf{w}^H (\mathbf{h}_1 s_1 + \mathbf{h}_2 s_2) \\
 &= \frac{\hat{\mathbf{h}}_1^H \mathbf{K}_x^{-1}}{\hat{\mathbf{h}}_1^H \mathbf{K}_x^{-1} \hat{\mathbf{h}}_1} (\mathbf{h}_1 s_1 + \mathbf{h}_2 s_2) \\
 &= \frac{\hat{\mathbf{h}}_1^H \mathbf{K}_x^{-1} \mathbf{h}_1}{\hat{\mathbf{h}}_1^H \mathbf{K}_x^{-1} \hat{\mathbf{h}}_1} s_1 + \frac{\hat{\mathbf{h}}_1^H \mathbf{K}_x^{-1} \mathbf{h}_2}{\hat{\mathbf{h}}_1^H \mathbf{K}_x^{-1} \hat{\mathbf{h}}_1} s_2.
 \end{aligned} \tag{4.21}$$

Im folgenden soll die Eigenstruktur der Kovarianzmatrix genauer untersucht werden, weil man die in (4.21) benötigte Inverse von \mathbf{K}_x über die Spektralzerlegung leicht beschreiben kann:

$$\mathbf{K}_x^{-1} = \sum \frac{1}{\lambda_i} \mathbf{u}_i \mathbf{u}_i^H. \tag{4.22}$$

Die Kovarianzmatrix der Störgeräusche läßt sich mit den in Kap. 4.1 gemachten Annahmen über isotropes, räumlich unkorreliertes Rauschen in zwei Teile aufspalten:

$$\mathbf{K}_x = \sigma_{s_2}^2 \mathbf{h}_2 \mathbf{h}_2^H + \sigma_n^2 \mathbf{I} = \mathbf{K}_{xs} + \mathbf{K}_n. \tag{4.23}$$

Zunächst soll ein Eigenvektor von \mathbf{K}_{xs} gefunden werden.

Lemma 1. *Ein Eigenvektor von \mathbf{K}_{xs} ist $\mathbf{u}_1 = \frac{\mathbf{h}_2}{|\mathbf{h}_2|}$.*

Beweis.

$$\begin{aligned}
 \sigma_s \mathbf{h}_2 \mathbf{h}_2^H \mathbf{u}_1 &= \lambda_1 \mathbf{u}_1 \\
 \sigma_s \mathbf{h}_2 \mathbf{h}_2^H \frac{\mathbf{h}_2}{|\mathbf{h}_2|} &= \lambda_1 \frac{\mathbf{h}_2}{|\mathbf{h}_2|} \\
 \sigma_s \mathbf{h}_2 |\mathbf{h}_2|^2 &= \lambda_1 \mathbf{h}_2
 \end{aligned}$$

Damit ist \mathbf{u}_1 tatsächlich ein Eigenvektor, und der zugehörige Eigenwert ist $\lambda_{s_1} = \sigma_s |\mathbf{h}_2|^2 = \sigma_s$. q.e.d.

\mathbf{K}_{xs} hat keine weiteren von Null verschiedenen Eigenwerte, weil sie den Rang eins hat. Weiterhin ist sie als Kovarianzmatrix eine hermitesche Matrix, so daß ihr Kern¹¹ senkrecht (unitär) zum Bild¹² steht. Sei $\mathbf{u}_2 = \mathbf{h}_{2\perp}/|\mathbf{h}_2|$ ein Vektor aus dem Kern von \mathbf{K}_{xs} und sein „Eigenwert“ $\lambda_{s_2} = 0$.

Lemma 2. *\mathbf{K}_x hat die Eigenwerte $\lambda_1 = \sigma_{s_2}^2 + \sigma_n^2$ und $\lambda_2 = \sigma_n^2$. Die zugehörigen Eigenvektoren sind $\mathbf{u}_1 = \mathbf{h}_2/|\mathbf{h}_2|$ und $\mathbf{u}_2 = \mathbf{h}_{2\perp}/|\mathbf{h}_2|$ mit $\mathbf{h}_{2\perp}$ unitär zu \mathbf{h}_2 .*

¹¹Kern(\mathbf{A}) = $\{\mathbf{x} \in \mathbb{R}^N : \mathbf{0} = \mathbf{A}\mathbf{x}\}$, engl. nullspace.

¹²Bild(\mathbf{A}) = $\{\mathbf{y} \in \mathbb{R}^N : \mathbf{y} = \mathbf{A}\mathbf{x} \text{ für ein beliebiges } \mathbf{x} \in \mathbb{R}^N\}$, engl. range.

Beweis. Jeder Vektor ist ein Eigenvektor einer Einheitsmatrix, also auch \mathbf{u}_1 und \mathbf{u}_2 . Ihre Eigenwerte sind eins, so daß sich folgende Spektralzerlegung von $\mathbf{K} = \mathbf{U}^H \mathbf{\Lambda} \mathbf{U}$ ergibt:

$$\begin{aligned} \mathbf{K}_x &= \begin{bmatrix} \frac{\mathbf{h}_2^H}{|\mathbf{h}_2|} \\ \frac{\mathbf{h}_{2\perp}^H}{|\mathbf{h}_2|} \end{bmatrix} \left(\begin{bmatrix} \sigma_s^2 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \sigma_n^2 & 0 \\ 0 & \sigma_n^2 \end{bmatrix} \right) \begin{bmatrix} \frac{\mathbf{h}_2}{|\mathbf{h}_2|} & \frac{\mathbf{h}_{2\perp}}{|\mathbf{h}_2|} \\ \frac{\mathbf{h}_2}{|\mathbf{h}_2|} & \frac{\mathbf{h}_{2\perp}}{|\mathbf{h}_2|} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\mathbf{h}_2^H}{|\mathbf{h}_2|} \\ \frac{\mathbf{h}_{2\perp}^H}{|\mathbf{h}_2|} \end{bmatrix} \begin{bmatrix} \sigma_s^2 + \sigma_n^2 & 0 \\ 0 & \sigma_n^2 \end{bmatrix} \begin{bmatrix} \frac{\mathbf{h}_2}{|\mathbf{h}_2|} & \frac{\mathbf{h}_{2\perp}}{|\mathbf{h}_2|} \\ \frac{\mathbf{h}_2}{|\mathbf{h}_2|} & \frac{\mathbf{h}_{2\perp}}{|\mathbf{h}_2|} \end{bmatrix}. \end{aligned}$$

Folglich sind \mathbf{u}_1 und \mathbf{u}_2 Eigenvektoren von \mathbf{K}_x . Ihre Eigenwerte sind $\lambda_1 = \sigma_s^2 + \sigma_n^2$ und $\lambda_2 = \sigma_n^2$. q.e.d.

Damit kann man (4.21) leicht analysieren. Der erste Term zeigt, daß das Nutzsinal den Beamformer dann ungehindert passiert, wenn $\hat{\mathbf{h}}_1^H = \mathbf{h}_1^H$ ist, denn dann ist der Zähler gleich dem Nenner. Eine ungenaue Schätzung führt zu einer frequenzabhängigen Modifikation der Amplitude von y und folglich zu einer Verzerrung des breitbandigen Signals $\hat{s}^b = y^b$.

Der zweite Term zeigt den im Ausgangssignal verbleibenden Rest des Störgeräusches s_2 und gibt somit die Qualität der Störunterdrückung an. Er soll nun analysiert werden. Es gilt:

$$\begin{aligned} \hat{\mathbf{h}}_1^H \mathbf{K}_x^{-1} \mathbf{h}_2 &= \hat{\mathbf{h}}_1^H \left(\frac{\mathbf{u}_1 \mathbf{u}_1^H}{\lambda_1} + \frac{\mathbf{u}_2 \mathbf{u}_2^H}{\lambda_2} \right) \mathbf{h}_2 \\ &= \hat{\mathbf{h}}_1^H \left(\frac{\mathbf{h}_2 \mathbf{h}_2^H}{(\sigma_{s_2}^2 + \sigma_n^2)} + \frac{\mathbf{h}_{2\perp} \mathbf{h}_{2\perp}^H}{\sigma_n^2} \right) \mathbf{h}_2 \\ &= \hat{\mathbf{h}}_1^H \frac{\mathbf{h}_{2\perp} \mathbf{h}_{2\perp}^H}{\sigma_n^2} \mathbf{h}_2 = \frac{\hat{\mathbf{h}}_1^H \mathbf{h}_{2\perp}}{\sigma_n^2} \mathbf{h}_{2\perp}^H \mathbf{h}_2 = 0 \quad (4.24) \\ \text{iff} \quad &\sigma_{s_2}^2 \gg \sigma_n^2. \end{aligned}$$

Das letzte Gleichheitszeichen zeigt, daß $\hat{\mathbf{h}}_1^H \mathbf{K}^{-1}$ orthogonal (hier unitär) zu der Störung $\mathbf{h}_2 s_2$ steht und somit das Orthogonalitätsprinzip (Therrien (1992)) erfüllt. Damit ist dieser Schätzer optimal im Sinne einer MMSE-Fehlerminimierung.

Bedeutsamer ist jedoch, daß die Störunterdrückung keine Abhängigkeit von $\hat{\mathbf{h}}_1$ besitzt. Die Orthogonalität in (4.24) ist immer gewährleistet, solange die Bedingung eingehalten ist, daß das isotrope Rauschen im Verhältnis zum kohärenten Störsignal hinreichend gering ist. Die Vorteil wird jedoch durch die Notwendigkeit „erkauft“, \mathbf{K}_x ohne die „Störung“ durch das Nutzsinal schätzen zu müssen. Die Möglichkeit allerdings nicht immer, so daß oft auch eine Signalrekonstruktion mit dem MV-Beamformer erfolgen muß.

4.4.3 Konsequenzen für den MV-Beamformer

Analog zum vorhergehenden Kapitel wird zunächst eine Gleichung aufgestellt, die die Beiträge des Nutz- und Störsignals auf die Ausgangssignale beschreibt. Dann

werden die einzelnen Teile mit Hilfe der Eigenstruktur der Kovarianzmatrix der gemessenen Signale untersucht.

Das Ausgangssignal ergibt sich zu

$$\begin{aligned}
 y &= \mathbf{w}^H (\mathbf{h}_1 s_1 + \mathbf{h}_2 s_2) \\
 &= \frac{\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1}}{\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \hat{\mathbf{h}}_1} (\mathbf{h}_1 s_1 + \mathbf{h}_2 s_2) \\
 &= \frac{\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \mathbf{h}_1}{\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \hat{\mathbf{h}}_1} s_1 + \frac{\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \mathbf{h}_2}{\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \hat{\mathbf{h}}_1} s_2 \\
 &= p_1 s_1 + p_2 s_2.
 \end{aligned} \tag{4.25}$$

Damit beschreiben p_1 und p_2 die Einflüsse des Nutz- bzw. des Störsignals auf den Ausgang. Um sie zu analysieren, wird die Eigenstruktur von \mathbf{R}_x untersucht. \mathbf{R}_x setzt sich aus drei Teilen zusammen:

$$\mathbf{R}_x = \sigma_{s_1}^2 \mathbf{h}_1 \mathbf{h}_1^H + \sigma_{s_2}^2 \mathbf{h}_2 \mathbf{h}_2^H + \sigma_n^2 \mathbf{I}. \tag{4.26}$$

Da wieder vorausgesetzt wird, daß $\sigma_{s_2}^2 \gg \sigma_n^2$, kann man den letzten Summanden vernachlässigen. Weiterhin gelten die im vorigen Kapitel (Kap. 4.4.2) gemachten Vereinfachungen.

Lemma 3. Die Kovarianzmatrix \mathbf{R}_x hat die Eigenvektoren und -werte

$$\mathbf{u}_1 = \frac{\mathbf{h}_1 + \kappa \mathbf{h}_2}{|\mathbf{h}_1 + \kappa \mathbf{h}_2|} \quad \lambda_1 = \sigma_s^2 (|\mathbf{h}_1| + |\mathbf{h}_1^H \mathbf{h}_2|) \tag{4.27}$$

$$\mathbf{u}_2 = \frac{\mathbf{h}_1 - \kappa \mathbf{h}_2}{|\mathbf{h}_1 - \kappa \mathbf{h}_2|} \quad \lambda_2 = \sigma_s^2 (|\mathbf{h}_1| - |\mathbf{h}_1^H \mathbf{h}_2|) \tag{4.28}$$

mit $\kappa = \mathbf{h}_2^H \mathbf{h}_1 / |\mathbf{h}_2^H \mathbf{h}_1|$

Beweis. Ein potentieller Eigenvektor \mathbf{u} läßt sich als Linearkombination von \mathbf{h}_1 und \mathbf{h}_2 darstellen: $\mathbf{u} = \alpha \mathbf{h}_1 + \beta \mathbf{h}_2$. Weiterhin wird $b = |\mathbf{h}_1|^2 = \mathbf{h}_1^H \mathbf{h}_1 = |\mathbf{h}_2|^2$ und $d = \mathbf{h}_1^H \mathbf{h}_2$ gesetzt. Bedeutsam ist, daß das Skalarprodukt für komplexwertige Vektoren im Gegensatz zu reellwertigen nicht kommutativ ist: $\mathbf{a}^H \mathbf{b} = (\mathbf{b}^H \mathbf{a})^*$. Damit ist $d^* = \mathbf{h}_2^H \mathbf{h}_1$.

$$\begin{aligned}
 \lambda(\alpha \mathbf{h}_1 + \beta \mathbf{h}_2) &= (\sigma_{s_1}^2 \mathbf{h}_1 \mathbf{h}_1^H + \sigma_{s_2}^2 \mathbf{h}_2 \mathbf{h}_2^H)(\alpha \mathbf{h}_1 + \beta \mathbf{h}_2) \\
 \lambda \alpha \mathbf{h}_1 + \lambda \beta \mathbf{h}_2 &= \alpha \sigma_{s_1}^2 \mathbf{h}_1 \mathbf{h}_1^H \mathbf{h}_1 + \alpha \sigma_{s_2}^2 \mathbf{h}_2 \mathbf{h}_2^H \mathbf{h}_1 \\
 &\quad + \beta \sigma_{s_1}^2 \mathbf{h}_1 \mathbf{h}_1^H \mathbf{h}_2 + \beta \sigma_{s_2}^2 \mathbf{h}_2 \mathbf{h}_2^H \mathbf{h}_2 \\
 &= \alpha \sigma_s^2 b \mathbf{h}_1 + \alpha \sigma_s^2 d^* \mathbf{h}_2 + \beta \sigma_s^2 d \mathbf{h}_1 + \beta \sigma_s^2 b \mathbf{h}_2 \\
 &= \sigma_s^2 (\alpha b + \beta d) \mathbf{h}_1 + \sigma_s^2 (\alpha d^* + \beta b) \mathbf{h}_2
 \end{aligned}$$

Durch Koeffizientenvergleich kommt man auf zwei Gleichungen:

$$\begin{aligned}
 \sigma_s^2 (\alpha b + \beta d) &= \lambda \alpha \\
 \sigma_s^2 (\alpha d^* + \beta b) &= \lambda \beta.
 \end{aligned}$$

Setzt man $\alpha = 1$, ergibt sich

$$\beta = \frac{\lambda - \sigma_s^2 b}{\sigma_s^2 d} \quad 0 = (\sigma_s^2)^2 d d^* + \frac{(\lambda - \sigma_s^2 b)}{\sigma_s^2 d} (\sigma_s^2 b - \lambda).$$

Die Lösung der quadratischen Gleichung ist $\lambda_{1/2} = \sigma_s^2 (b \pm |d|)$, damit sind die Eigenwerte

$$\lambda_1 = \sigma_s^2 (|\mathbf{h}_1|^2 + |\mathbf{h}_1^H \mathbf{h}_2|) \quad \lambda_2 = \sigma_s^2 (|\mathbf{h}_1|^2 - |\mathbf{h}_1^H \mathbf{h}_2|).$$

β ergibt sich durch Einsetzen

$$\beta = \frac{\lambda - \sigma_s^2 b}{\sigma_s^2 d} = \frac{\sigma_s^2 (b \pm |d|) - \sigma_s^2 b}{\sigma_s^2 d} = \frac{\pm |d|}{d}.$$

Setzt man $\kappa = |d|/d = d^*/|d|$, ergeben sich die Eigenvektoren nach einer Normierung zu

$$\mathbf{u}_1 = \frac{\mathbf{h}_1 + \kappa \mathbf{h}_2}{|\mathbf{h}_1 + \kappa \mathbf{h}_2|} \quad \mathbf{u}_1 = \frac{\mathbf{h}_1 - \kappa \mathbf{h}_2}{|\mathbf{h}_1 - \kappa \mathbf{h}_2|}.$$

q.e.d.

Korollar 1. *Aus diesem Lemma folgen die Eigenvektoren und -werte für eine reellwertige Kovarianzmatrix als Spezialfall.*

$$\mathbf{u}_1 = \frac{\mathbf{h}_1 + \mathbf{h}_2}{|\mathbf{h}_1 + \mathbf{h}_2|} \quad \lambda_1 = \sigma_s^2 \mathbf{h}_1^H (\mathbf{h}_1 + \mathbf{h}_2) \quad (4.29)$$

$$\mathbf{u}_2 = \frac{\mathbf{h}_1 - \mathbf{h}_2}{|\mathbf{h}_1 - \mathbf{h}_2|} \quad \lambda_2 = \sigma_s^2 \mathbf{h}_1^H (\mathbf{h}_1 - \mathbf{h}_2). \quad (4.30)$$

Mit diesem Wissen können in (4.25) die Beiträge p_1 und p_2 des Nutz- und Störsignals zum Ausgangssignal bestimmt werden. Zunächst soll mit dem Zähler von p_2 begonnen werden:

$$\begin{aligned} \hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \mathbf{h}_2 &= \hat{\mathbf{h}}_1^H \left(\frac{\mathbf{u}_1 \mathbf{u}_1^H}{\lambda_1} + \frac{\mathbf{u}_2 \mathbf{u}_2^H}{\lambda_2} \right) \mathbf{h}_2 \\ &= \frac{\hat{\mathbf{h}}_1^H \mathbf{u}_1 \mathbf{u}_1^H \mathbf{h}_2}{\lambda_1} + \frac{\hat{\mathbf{h}}_1^H \mathbf{u}_2 \mathbf{u}_2^H \mathbf{h}_2}{\lambda_2}. \end{aligned}$$

Die oben ermittelten Eigenwerte und -vektoren werden in den ersten Summanden eingesetzt:

$$\frac{\hat{\mathbf{h}}_1^H \mathbf{u}_1 \mathbf{u}_1^H \mathbf{h}_2}{\lambda_1} = \frac{\hat{\mathbf{h}}_1^H (\mathbf{h}_1 + \kappa \mathbf{h}_2) (\mathbf{h}_1 + \kappa \mathbf{h}_2)^H \mathbf{h}_2}{\sigma_s^2 (\mathbf{h}_1 + \kappa \mathbf{h}_2)^H (\mathbf{h}_1 + \kappa \mathbf{h}_2) (|\mathbf{h}|^2 + |\mathbf{h}_1^H \mathbf{h}_2|)}.$$

Es gilt:

$$(\mathbf{h}_1 + \kappa \mathbf{h}_2)^H \mathbf{h}_2 = \mathbf{h}_1^H \mathbf{h}_2 + \kappa^* |\mathbf{h}|^2 = \kappa^* |\mathbf{h}_1^H \mathbf{h}_2| + \kappa^* |\mathbf{h}|^2.$$

Damit sind die letzten Terme skalar und identisch und lassen sich kürzen:

$$\frac{\hat{\mathbf{h}}_1^H \mathbf{u}_1 \mathbf{u}_1^H \mathbf{h}_2}{\lambda_1} = \frac{\kappa^* \hat{\mathbf{h}}_1^H (\mathbf{h}_1 + \kappa \mathbf{h}_2)}{\sigma_s^2 (\mathbf{h}_1 + \kappa \mathbf{h}_2)^H (\mathbf{h}_1 + \kappa \mathbf{h}_2)}.$$

Weiterhin gilt:

$$(\mathbf{h}_1 + \kappa \mathbf{h}_2)^H (\mathbf{h}_1 + \kappa \mathbf{h}_2) = |\mathbf{h}|^2 + \kappa^* \mathbf{h}_2^H \mathbf{h}_1 + \kappa \mathbf{h}_1^H \mathbf{h}_2 + |\mathbf{h}|^2 = 2 \cdot |\mathbf{h}|^2 + 2 \cdot |\mathbf{h}_1^H \mathbf{h}_2|.$$

Damit ergibt sich

$$\frac{\hat{\mathbf{h}}_1^H \mathbf{u}_1 \mathbf{u}_1^H \mathbf{h}_2}{\lambda_1} = \frac{\kappa^* \hat{\mathbf{h}}_1^H \mathbf{h}_1 + \hat{\mathbf{h}}_1^H \mathbf{h}_2}{2\sigma_s^2 (|\mathbf{h}|^2 + |\mathbf{h}_1^H \mathbf{h}_2|)}.$$

Analog gilt für den zweiten Term

$$\frac{\hat{\mathbf{h}}_1^H \mathbf{u}_2 \mathbf{u}_2^H \mathbf{h}_2}{\lambda_1} = -\frac{\kappa^* \hat{\mathbf{h}}_1^H \mathbf{h}_1 - \hat{\mathbf{h}}_1^H \mathbf{h}_2}{2\sigma_s^2 (|\mathbf{h}|^2 - |\mathbf{h}_1^H \mathbf{h}_2|)}$$

$$\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \mathbf{h}_2 = \frac{\kappa^* \hat{\mathbf{h}}_1^H \mathbf{h}_1 + \hat{\mathbf{h}}_1^H \mathbf{h}_2}{2\sigma_s^2 (|\mathbf{h}|^2 + |\mathbf{h}_1^H \mathbf{h}_2|)} - \frac{\kappa^* \hat{\mathbf{h}}_1^H \mathbf{h}_1 - \hat{\mathbf{h}}_1^H \mathbf{h}_2}{2\sigma_s^2 (|\mathbf{h}|^2 - |\mathbf{h}_1^H \mathbf{h}_2|)}.$$

Mit den Definition aus (4.20) kann man die geometrische Bedeutung des Skalarprodukts ausnutzen:

$$\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \mathbf{h}_2 = \frac{\cos \mu + \cos(\alpha - \mu)}{2\sigma_s^2 \cdot (1 + \cos \alpha)} - \frac{\cos \mu - \cos(\alpha - \mu)}{2\sigma_s^2 \cdot (1 - \cos \alpha)}.$$

Weitere Vereinfachungen ergeben sich mit den Additionstheoremen:

$$\begin{aligned} \hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \mathbf{h}_2 &= \frac{\cos \mu + \cos(\alpha - \mu)}{2\sigma_s^2 \cdot (1 + \cos \alpha)} + \frac{\cos \mu - \cos(\alpha - \mu)}{2\sigma_s^2 \cdot (1 - \cos \alpha)} \\ &= \frac{\cos \mu + \cos \alpha \cos \mu + \sin \alpha \sin \mu}{2\sigma_s^2 \cdot (1 + \cos \alpha)} \\ &\quad - \frac{\cos \mu - \cos \alpha \cos \mu - \sin \alpha \sin \mu}{2\sigma_s^2 \cdot (1 - \cos \alpha)} \\ &= \frac{\cos \mu(1 + \cos \alpha)}{2\sigma_s^2 \cdot (1 + \cos \alpha)} - \frac{\sin \alpha \sin \mu}{2\sigma_s^2 \cdot (1 + \cos \alpha)} \\ &\quad - \frac{\cos \mu(1 - \cos \alpha)}{2\sigma_s^2 \cdot (1 - \cos \alpha)} + \frac{\sin \alpha \sin \mu}{2\sigma_s^2 \cdot (1 - \cos \alpha)} \\ &= \frac{\sin \alpha \sin \mu}{2\sigma_s^2} \left(\frac{1}{1 + \cos \alpha} + \frac{1}{1 - \cos \alpha} \right) \\ &= \frac{\sin \alpha \sin \mu}{2\sigma_s^2} \left(\frac{1 - \cos \alpha + (1 + \cos \alpha)}{(1 + \cos \alpha)(1 - \cos \alpha)} \right) \\ &= \frac{\sin \alpha \sin \mu}{\sigma_s^2 (1 - \cos^2 \alpha)} = \frac{\sin \alpha \sin \mu}{\sigma_s^2 \sin^2 \alpha} \\ &= \frac{\sin \mu}{\sigma_s^2 \sin \alpha}. \end{aligned} \tag{4.31}$$

Analog kann man mit dem Zähler von p_1 verfahren:

$$\begin{aligned}
\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \mathbf{h}_1 &= \frac{\hat{\mathbf{h}}_1^H \mathbf{h}_1 + \hat{\mathbf{h}}_1^H \mathbf{h}_2}{2\sigma_s^2 \cdot (\mathbf{h}_1^H \mathbf{h}_1 + \mathbf{h}_1^H \mathbf{h}_2)} + \frac{\hat{\mathbf{h}}_1^H \mathbf{h}_1 - \hat{\mathbf{h}}_1^H \mathbf{h}_2}{2\sigma_s^2 \cdot (\mathbf{h}_1^H \mathbf{h}_1 - \mathbf{h}_1^H \mathbf{h}_2)} \\
&= \frac{\cos \mu + \cos \alpha \cos \mu + \sin \alpha \sin \mu}{2\sigma_s^2 \cdot (1 + \cos \alpha)} + \frac{\cos \mu - \cos \alpha \cos \mu - \sin \alpha \sin \mu}{2\sigma_s^2 \cdot (1 - \cos \alpha)} \\
&= \frac{\cos \mu(1 + \cos \alpha)}{2\sigma_s^2 \cdot (1 + \cos \alpha)} + \frac{\sin \alpha \sin \mu}{2\sigma_s^2 \cdot (1 + \cos \alpha)} \\
&\quad - \frac{\cos \mu(1 - \cos \alpha)}{2\sigma_s^2 \cdot (1 - \cos \alpha)} - \frac{\sin \alpha \sin \mu}{2\sigma_s^2 \cdot (1 - \cos \alpha)} \\
&= \frac{\cos \mu}{\sigma_s^2} + \\
&\quad \frac{\sin \alpha \sin \mu}{2\sigma_s^2} \left(\frac{1}{1 + \cos \alpha} - \frac{1}{1 - \cos \alpha} \right) \\
&= \frac{\cos \mu}{\sigma_s^2} + \\
&\quad \frac{\sin \alpha \sin \mu}{2\sigma_s^2} \left(\frac{1 - \cos \alpha - (1 + \cos \alpha)}{(1 + \cos \alpha)(1 - \cos \alpha)} \right) \\
&= \frac{\cos \mu}{\sigma_s^2} - \frac{\sin \alpha \sin \mu \cos \alpha}{\sigma_s^2(1 - \cos \alpha^2)} \\
&= \frac{\cos \mu - \cot \alpha \sin \mu}{\sigma_s^2}. \tag{4.32}
\end{aligned}$$

Schließlich ergibt sich für den Nenner von p_1 und p_2 :

$$\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \hat{\mathbf{h}}_1 = \frac{1}{\sigma_s^2} \left(\frac{\cos \mu + \cos(\alpha - \mu)}{2 \cdot (1 + \cos \alpha)} \right)^2 + \frac{1}{\sigma_s^2} \left(\frac{\cos \mu - \cos(\alpha - \mu)}{2 \cdot (1 - \cos \alpha)} \right)^2. \tag{4.33}$$

Wenn (und nur wenn) $\hat{\mathbf{h}}_1$ korrekt geschätzt ist, ist der Steeringvektor unitär zum Störsignal und damit das Orthogonalitätsprinzip erfüllt, d.h. das Störsignal wird optimal unterdrückt. Das Nutzsignal wird mit dem Faktor 1 für alle Frequenzen multipliziert, es wird also nicht abgeschwächt oder verstärkt, sondern unverzerrt durchgelassen:

$$p_1 = \mathbf{w}_{\text{MV}}^H \hat{\mathbf{h}}_1 = \frac{\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \mathbf{h}_1}{\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \mathbf{h}_1} = \frac{\frac{1}{2\sigma^2} - \frac{1}{2\sigma^2}}{\sigma^2} = 1 \tag{4.34}$$

$$p_2 = \mathbf{w}_{\text{MV}}^H \hat{\mathbf{h}}_2 = \frac{\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \mathbf{h}_2}{\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \mathbf{h}_1} = \frac{\frac{1}{2\sigma^2} + \frac{1}{2\sigma^2}}{\sigma^2} = 0. \tag{4.35}$$

Wird $\hat{\mathbf{h}}_1$ hingegen aus den am Anfang genannten Gründen nicht genau geschätzt, kommt es zu einem drastischen Einbruch der Leistungsfähigkeit. Abb. 4.9 zeigt die Performance (d.h. die Verstärkung, die das Nutz- und Störsignal erfahren) in Abhängigkeit von μ für verschiedene α . Mit steigendem μ wird das Nutzsignal erheblich gedämpft. Da die Dämpfung frequenzabhängig ist, führt das zu einer zusätzlichen

Verzerrung des Nutzsignals. Weiterhin wird das Orthogonalitätsprinzip in zunehmendem Maße verletzt, so daß das Störsignal sogar das Nutzsignal übersteigt. Folglich ist ein solcher Beamformer unter suboptimalen Bedingungen nur eingeschränkt einsatzfähig.

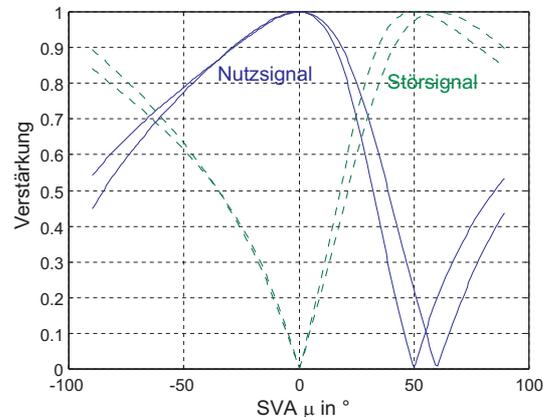


Abb. 4.9: Verstärkung des Nutzsignals (durchgezogen) und des Störsignals (gestrichelt) in Abhängigkeit vom Schätzfehlerwinkel. Die Werte von α betragen 50° und 60° .

4.4.4 Möglichkeiten für einen robusten Beamformer

Um einen robusten Beamformer zu erhalten, sind verschiedene Maßnahmen vorgeschlagen worden. Zunächst wird fast ausschließlich¹³ der ML-Beamformer eingesetzt, dessen Störunterdrückung deutlich robuster ist als die des MV-Beamformers. In der Anwendung unterscheiden sich die beiden Beamformer nur dadurch, daß der ML-Beamformer lediglich in den Signalpausen adaptiert wird¹⁴.

Die Methode, die in der Sprachverarbeitung üblich ist, kann für die Maschinendiagnose nicht eingesetzt werden, weil oft ein großer zeitlicher Abstand zwischen der Aufnahme der zu klassifizierenden Signale und der Möglichkeit, die Kovarianzmatrix zu schätzen, herrscht. Die Nebengeräusche sind zu instationär, um zu beiden Zeitpunkten die gleichen statistischen Parameter zu haben. Weiterhin treten auch hier Verzerrungen des Nutzsignals auf.

Darum wurden weitere Randbedingungen vorgeschlagen, um einen robusten Beamformer zu garantieren. Die ersten Vorschläge waren quadratische Randbedingungen, die von Nunn (1983) vorgeschlagen wurden. Die heute üblichen Methoden werden unter dem Begriff „Maximally flat spatial power response“ zusammengefaßt (vgl. Buckley und Griffith (1986), Shutao und Thng (2001) oder Hoshuyama und Sugiyama (2001), die auch viele weitere Referenzen angeben). Sie besitzen den Vorteil

¹³zumindest in Japan.

¹⁴Das führt nicht selten zu begrifflichen Komplikationen. Gelegentlich wird von einem Minimum Variance Beamformer gesprochen, d.h. von einem Beamformer, der direkt auf Frost zurückgeht, obwohl die Adaption nur in den Signalpausen erfolgt, wenn das Nutzsinal nicht präsent ist und die komplette Kovarianzmatrix auch nicht geschätzt werden kann.

linearer Randbedingungen. Die Gemeinsamkeit all dieser Methoden liegt darin, daß sie die Steilheit des Beampatterns und somit den Einfluß einer Fehlschätzung reduzieren. Da die Steilheit des Beampatterns die Störunterdrückung bestimmt, heißt das, daß man sich die höhere Robustheit durch eine geringe Störunterdrückung erkaufen muß.

Kapitel 5

Blinde Trennung akustischer Mischungen

Die Blinde Quellentrennung (BSS) ist ein neuer Ansatz der mehrkanaligen Signalverarbeitung, die im Gegensatz zu den bisher beschriebenen Beamformern, keine Information über die Positionen der Quellen, die spektrale Konstruktion oder die Art der Veränderung im Raum benötigt. Die theoretische Basis ist die Independent Component Analysis (ICA).

Deren Grundidee besteht darin, statistisch unabhängige Signale zu rekonstruieren. Wenn statistisch unabhängige Signale (Quellensignale) linear überlagert werden, sind die überlagerten Signale (Observationen) linear und damit auch statistisch abhängig. Die einzige Möglichkeit, eine statistische Unabhängigkeit zu erreichen, ist, die Überlagerung rückgängig zu machen. Damit ist die statistische Unabhängigkeit eine notwendige und hinreichende Bedingung dafür, daß die geschätzten Signale den Quellensignalen entsprechen.

Die statistische Abhängigkeit der Signale wird verschieden gemessen. Sie basiert auf der räumlichen Statistik (d.h. der Statistik zwischen den Kanälen) höherer Ordnung, der räumlich-temporalen Statistik zweiter Ordnung, dem Grad der sparsamen Kodierbarkeit (sparseness) oder der Beschreibungslänge (Maximal descriptions length). Welche der Methoden zu wählen ist, richtet sich nach der Art der Signale¹.

Weiterhin müssen Wege gefunden werden, wie die Überlagerung der Signale rückgängig gemacht werden kann. Im einfachsten Fall, in dem die Observationen durch gewichtete Summen aus den Quellensignalen entstanden sind (skalare Mischungen), kann man die Überlagerung ebenfalls durch eine gewichtete Summe rückgängig machen. Die Gewichte werden so eingestellt, daß die Unabhängigkeit der so geschätzten Signale maximal ist.

Die in dieser Arbeit zu behandelnden Überlagerungen akustischer Signale beinhalten Echos und Zeitverzögerungen, d.h. gefilterte Signale werden überlagert (konvolute Mischungen). Um die Überlagerungen rückgängig zu machen, müssen

¹Damit ist die BSS streng genommen nicht blind, weil die Methode zur Beschreibung der statistischen Abhängigkeit nur in Abhängigkeit von den Signalen gewählt werden kann. Der Name wird dennoch beibehalten, weil die Informationen über die Signale nur allgemeiner Natur sein müssen.

die inversen Filter gefunden werden. Damit ist die Trennung akustischer Mischungen erheblich aufwendiger als die Trennung skalarer Mischungen, weil die Anzahl der zu schätzenden Filterkoeffizienten deutlich höher ist.

Zunächst wird in diesem Kapitel die Grundidee für die blinde Trennung konvolutiver Mischungen mit der ICA vorgestellt. Im folgenden werden die wichtigsten Methoden beschrieben. Im Abschluß werden Lösungsvorschläge zu zwei zentralen Problemen, dem Skalierungs- und dem Permutationsproblem, diskutiert.

5.1 Idee der blinden Quellentrennung

5.1.1 Darstellung der statistischen Unabhängigkeit

Um die statistische Abhängigkeit von Zufallsvariablen und -prozessen zu minimieren, muß man sie formalisieren und quantifizieren².

Zwei Zufallsvariablen x und y sind statistisch unabhängig, wenn die Kenntnis der einen Variablen keine Aussage über den Zustand der anderen Variablen zuläßt. Für Signale, die als stochastische Prozesse aufgefaßt werden können, ist das (wenigstens dann) garantiert, wenn sie verschiedene physikalische Ursachen haben, die in keiner Beziehung zueinander stehen. In den weiteren Betrachtungen sei das als gegeben angenommen.

Mathematisch werden die Variablen als statistisch unabhängig definiert, wenn (und nur wenn) ihre Verbundwahrscheinlichkeitsdichte p_{xy} das Produkt der Randverteilungen $p_y \cdot p_x$ ist (Schwarz (1988)).

$$p_{xy}(x, y) = p_y(y) \cdot p_x(x) \quad (5.1)$$

Die bedingte Wahrscheinlichkeit $p(x|y)$ ist gleich $p(x)$. Für einen N -dimensionalen Zufallsvektor kann (5.1) in folgende Gleichung überführt werden:

$$p_{\mathbf{x}}(\mathbf{x}) = p_{x_1}(x_1) \cdot \dots \cdot p_{x_N}(x_N) = \prod_{n=1}^N p_{x_n}(x_n). \quad (5.2)$$

Äquivalent zu (5.1) ist eine weitere wichtige Eigenschaft statistisch unabhängiger Variablen für beliebige, absolut integrierbare Funktionen $g(x)$ und $h(y)$

$$E\{g(x)h(y)\} = E\{g(x)\}E\{h(y)\}, \quad (5.3)$$

d.h. alle nichtlinearen Abbildungen sind unkorreliert.

Gleichung (5.3) kann man für die Beschreibung der statistischen Abhängigkeit verwenden. Für die Maximierung der statistischen Unabhängigkeit werden dabei die Korrelationen der nichtlinearen Abbildungen der Variablen minimiert. Darum

²Für einen umfassenden Überblick sind die Lehrbücher von Cover und Thomas (1991), MacKay (2001), Hyvärinen *et al.* (2001) oder andere Lehrbücher über die Informationstheorie geeignet.

werden solche Verfahren als „nichtlineare Dekorrelation“ bezeichnet. Sie werden in Kap. 5.2 beschrieben. Da man jede beliebige nichtlineare Funktion durch ihre Taylorreihe ausdrücken kann, entstehen bei der Erwartungswertbildung Momente höherer Ordnung, so daß diese Verfahren implizit die Statistik höherer Ordnung benutzen.

Setzt man für $g(x)$ und $h(y)$ lineare Funktionen ein, ergibt sich als notwendige Bedingung, daß statistisch unabhängige Variablen unkorreliert sind. Die Zufallsvariablen, von denen entsprechend dem Signalmodell in Kap. 4.1 ausgegangen wird, sind als Schallsignale definitionsgemäß mittelwertfrei. Damit gilt

$$E\{xy\} = E\{x\}E\{y\} = 0. \quad (5.4)$$

Die Unkorreliertheit ist keine hinreichende Bedingung, weil (5.3) auch für alle anderen Funktionen gelten muß. Eine klassische Dekorrelation der Observationen ist also für die Schätzung der Quellensignale nicht hinreichend.

Weiterhin folgt daraus, daß die Zufallsvariablen nicht normalverteilt sein dürfen, wenn man die nichtlineare Dekorrelation zur Trennung benutzen will. Sind sie nämlich normalverteilt, ist die Wahrscheinlichkeitsdichte mit der Statistik zweiter Ordnung vollständig beschrieben und aus (5.3) folgen keine weiteren Bedingungen.

Handelt es sich jedoch bei den Zufallsvariablen um stochastische Prozesse, kann man auch mit der Statistik zweiter Ordnung hinreichende Bedingungen für die statistische Unabhängigkeit herleiten³. Für instationäre Prozesse muß (5.4) für jeden Zeitabschnitt gelten (siehe Weinstein *et al.* (1993), Parra und Spence (2000), Ikram und Morgan (2000)). Damit kann die Unkorreliertheit für alle Zeitabschnitte als hinreichende Bedingung für die Unabhängigkeit angesehen werden.

$$\forall t : E\{x(t)y(t)\} = 0 \quad (5.5)$$

Methoden auf dieser Basis werden als „Dekorrelation instationärer Signale“ (non-stationary decorrelation) bezeichnet und in Kap. 5.3 genauer beschrieben. Für ihre Anwendung ist die Instationarität der Prozesse nötig, sonst liefert (5.5) keine neuen Bedingungen. Dafür dürfen die Prozesse normalverteilt sein.

Weiterhin muß die Kreuzkovarianzfunktion für statistisch unabhängige Prozesse zu allen Zeitpunkten verschwinden, woraus sich ebenfalls eine hinreichende Bedingung für die Unabhängigkeit der Prozesse ergibt:

$$\forall \tau : E\{x(t)y(t + \tau)\} = r_{xy}(\tau) = 0. \quad (5.6)$$

Die Verfahren werden als „zeitverzögerte Dekorrelation“ (time-delayed decorrelation) bezeichnet und ebenfalls in Kap. 5.3 beschrieben. Die Voraussetzung für diese ist das

³Diese Bedingungen sind *nicht* äquivalent mit (5.1), weil sie eine Zeitkomponente enthalten. Sie sind lediglich analoge Eigenschaften statistisch unabhängiger Prozesse. Darum zählen einige Autoren Verfahren, die nach diesen Bedingungen optimieren, nicht zur ICA, wie Torkkola (2000), weil sie nicht die statistische Unabhängigkeit i.e.S. optimieren. Andere Autoren, wie Parra und Spence (2000), zählen sie zur ICA im weiteren Sinn, weil die Definitionen aus (5.5) und (5.6) ähnliche Eigenschaften haben: Sie gelten für Zufallsprozesse, die aus unterschiedlichen physikalischen Ursachen stammen, und gehen mit der linearen Überlagerung verloren.

Vorhandensein einer Autokovarianzfunktion (d.h. $r_{xx}(\tau \neq 0) \neq 0$), weil (5.6) sonst immer erfüllt ist. Das heißt, weißes Rauschen muß ausgeschlossen werden. Dafür dürfen die Prozesse normalverteilt und stationär sein.

Andere Wege, die statistische Abhängigkeit zu beschreiben, lassen sich daraus herleiten, daß unabhängige Prozesse eine minimale Kodierungslänge haben oder daß Transformationen existieren, in denen unabhängige Signale schwach besetzte Vektoren bzw. Matrizen bilden. Verfahren auf dieser Grundlage heißen „coding length minimization“ bzw. „sparseness maximization“ (Hyvärinen *et al.* (2001)).

5.1.2 Mischungs- und Entmischungsmodell für akustische Mischungen

Die ICA-basierten Verfahren sollen, ähnlich wie die Beamformer in Kap. 4, für die Rekonstruktion eines Nutzsignals aus einer akustischen Überlagerung verwendet werden. Dazu wird als Basis das Signalmodell benutzt, das in Kap. 4.1 eingeführt wurde.

Im allgemeinen geht die ICA von einem Mischungsmodell aus, bei dem M reellwertige, statistisch unabhängige Signale $\mathbf{s} = [s_1, \dots, s_M]^T$ durch $\mathbf{x} = \mathbf{H}\mathbf{s}$ gemischt werden, so daß die N Observationen $\mathbf{x} = [x_1, \dots, x_N]^T$ statistisch abhängig sind (skalare Mischung). Durch eine Multiplikation mit einer Matrix \mathbf{W} ($\mathbf{y} = \mathbf{W}\mathbf{x}$) versucht man nun statistisch unabhängige und damit überlagerungsfreie Signale, d.h. Rekonstruktionen der Quellensignale, zu erhalten. Dazu wurden zahlreiche funktionsfähige Algorithmen vorgeschlagen und evaluiert, die u.a. in den Lehrbüchern von Hyvärinen *et al.* (2001) oder Cichocki und Amari (2002) dokumentiert werden.

In der Akustik gilt dieses einfache Mischungsmodell nicht, weil die Mikrofone durch die Raumimpulsantwort verzerrte Signale empfangen (konvolutive Mischung), siehe Kap. 4.1. Dieses Problem umfaßt die Schätzung von erheblich mehr Parametern: Während bei der Entmischung skalarer Mischungen bei N Observationen „nur“ N^2 Parameter geschätzt werden müssen, sind es bei konvolutiven Mischungen wenigstens $N^2 \cdot K$ Parameter, wobei K die Länge der Raumimpulsantwort in Abtastschritten ist.

Die ersten Arbeiten zur Trennung konvolutiver Mischungen stammen von Weinstein *et al.* (1993), wobei damals noch nicht zwischen der blinden Quellentrennung i.e.S. und der mehrkanaligen blinden Entfaltung unterschieden wurde. Obwohl seitdem zahlreiche Algorithmen vorgeschlagen wurden, hat sich noch keine allgemein akzeptierte Lösung herausgebildet. Erste Übersichtsartikel sind von Torkkola (1999), Parra (1999) und Schobben *et al.* (1999) veröffentlicht worden, wobei Schobben *et al.* (1999) die Algorithmen auch vergleichen. Sie stellen jedoch den Stand von 1999 dar und dokumentieren neuere Algorithmen, die im Frequenzbereich arbeiten, noch nicht. Damit fehlen weitgehend eine einheitliche Nomenklatur und ein allgemein akzeptiertes Mischungsmodell.

Die Grundidee dieser Verfahren ist es, die Observationssignale adaptiv zu filtern und aufzuaddieren, wie es Abb. 5.1 zeigt. Die Filterkoeffizienten werden adaptiv so

eingestellt, daß die Ausgangssignale nach einer der Definitionen in Kap. 5.1.1 statistisch unabhängig sind⁴. Da jede lineare Überlagerung zu einer linearen und damit statistischen Abhängigkeit führt, ist die statistische Unabhängigkeit äquivalent mit der Überlagerungsfreiheit.

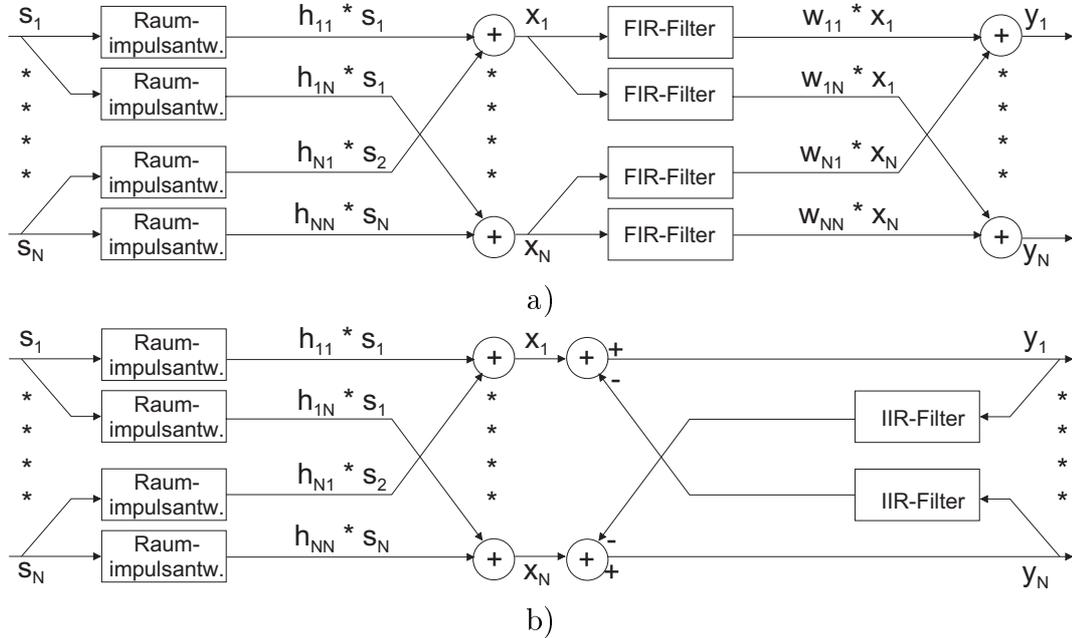


Abb. 5.1: Grundstruktur eines ICA-System zur Trennung akustischer Mischungen. a) FIR-System. b) rekursives (IIR-) System

Eine rückgekoppelte Realisierung mit IIR-Filtern, wie in Abb. 5.1b, ist zunächst naheliegend. Denn die Überlagerung mit Echos nach (4.2) kann man als FIR-Filter auffassen und die Inverse zu einem FIR-System ist ein IIR-System, das die Polstellen auf den Nullstellen des FIR-Systems hat. Die erste ICA-Realisierung für eine konvolute Mischung wurde von Torkkola (1996) mit einem IIR-Entmischungssystem realisiert.

Der entscheidende Nachteil dieser Realisierung besteht darin, daß die IIR-Filter instabil werden, wenn das Mischungssystem nicht-minimalphasig ist. Wie von Lee (1997) gezeigt, tritt eine nicht-minimalphasige akustische Filterung u.a. dann auf, wenn der Direktschall schwächer als die frühen Echos ist oder eine Totzeit in der Raumimpulsantwort liegt. Da das praktisch unvermeidbar ist, verbietet sich eine rückgekoppelte Realisierung für reale Probleme.

Bei dem Entmischungsmodell in Abb. 5.1a handelt es sich um ein FIR-Modell mit N^2K zu schätzenden Filterkoeffizienten, wie es unter anderem von Lee *et al.* (1997b) vorgeschlagen wurde

$$y_m^b(t) = \sum_{n=1}^N \sum_{k=0}^{K-1} w_{nmk}^b x_n^b(t-k). \quad (5.7)$$

⁴Das heißt, eine der Gleichungen (5.1), (5.3), (5.6) oder (5.5) muß erfüllt sein.

Lee *et al.* (1997b) benutzen ein zusätzliches Allpaß-Filter (reine Zeitverzögerung) und können damit ein „akausales“ Filter realisieren, das für die Inversion eines nicht-minimalphasigen Filters nötig ist. Damit ist dies das am häufigsten eingesetzte Modell.

Nachteilig ist die Tendenz dieser Systeme, die Signale auch zeitlich zu dekorrelieren. Bei der mit der blinden Trennung konvolutiver Mischungen eng verwandten mehrkanaligen blinden Entfaltung ist das erwünscht⁵, weil man von weißem Rauschen ausgeht und dessen Verzerrung durch den Raum rückgängig gemacht werden soll. Die Quellensignale in diesem Anwendungsfall, Maschinen- und Sprachsignale, weisen jedoch wie die meisten informationshaltigen Signale ein farbiges Leistungsdichtespektrum auf, d.h. sie sind zeitlich korreliert. Eine Dekorrelation ist also unerwünscht und man benötigt eine Vor- oder Nachbearbeitung, um eine Verflachung des Leistungsdichtespektrums zu verhindern. Eine Alternative ist die Formulierung der Unabhängigkeit im Frequenzbereich, wie sie im folgenden dargestellt wird.

5.1.3 ICA im Frequenzbereich

Für die Anwendung von (5.7) müssen nun die Filterkoeffizienten adaptiert werden. Die Filterkoeffizienten w_{nmk} werden adaptiert, bis die Ausgangssignale y_i^b, y_j^b die Bedingung

$$\forall i, j \neq i : E\{g(y_i^b)h(y_j^b)\} = 0 \quad (5.8)$$

oder eine andere der in Kap. 5.1.1 erarbeitete Bedingung erfüllen.

Verfahren, die diese Gleichung direkt zu erfüllen suchen, werden Verfahren im „Zeitbereich“ genannt. Sie besitzen kein Permutationsproblem, das im folgenden genauer besprochen wird. Jedoch haben sie das Problem, die Signale spektral zu weißer und bei langen Impulsantworten nicht zu den richtigen Lösungen zu konvergieren, weil mehrere hundert (oder tausend) von einander abhängige Parameter mit einer Gleichung adaptiert werden müssen.

Die Verfahren im Zeitbereich wurden von Lee *et al.* (1997a) eingeführt. Sie optimieren mit der Infomax-Bestimmungsgleichung (5.29) die Koeffizienten von FIR-Filtern unter der Verwendung der FIR-Filter- bzw. Polynomialgebra von Lambert (1996)⁶. Weitere Arbeiten auf dieser Basis erfolgten durch Lambert *et al.* (2001) oder Rahbar und Reilly (2001) mit sehr kurzen FIR-Filtern. Allerdings können diese Verfahren nicht für reale Raum-Impulsantworten verwendet werden. Wie von Araki *et al.* (2002) kürzlich gezeigt, können sie ab einer Filterlänge von 150 Taps nicht sinnvoll

⁵Beispiele für eine blinde Entfaltung finden sich vor allem in der Nachrichtenübertragung, bei der beispielsweise im Mobilfunk die Verzerrungen bestimmt und rückgängig gemacht werden sollen. Beispiele für Algorithmen, die ebenfalls die statistische Abhängigkeit minimieren, sind: Hua *et al.* (2001), die mit polynomialen Mischungsmatrizen arbeiten, wobei sie die Unterschiedlichkeit der Nullstellen als Kriterium der Unabhängigkeit benutzen; Luo *et al.* (2001), die die zeitverzögerte Dekorrelation anwenden, und andere.

⁶Diese Algebra verwendet Matrizen, deren Elemente FIR-Filter sind, sowie Polynommatrizen, die als z-Transformation der FIR-Filtermatrizen verwendet werden können. In diesem Raum werden dann die üblichen Matrizenoperationen, wie Eigenwerte, Inverse etc. definiert.

eingesetzt werden. In der Akustik sind sie lediglich für eine Subband-Verarbeitung (Araki *et al.* (2002)) oder für die Nachbearbeitung (Nishikawa *et al.* (2002)) sinnvoll.

Eine Ursache für die schlechte Konvergenz bei langen Raumimpulsantworten ist die Abhängigkeit zwischen den Filterkoeffizienten. Zu ihrer Verringerung ist eine Transformation in einen Raum sinnvoll, in dem die Koeffizienten nicht gekoppelt sind und damit unabhängig voneinander modifiziert werden können. Typischerweise ist das der Frequenzbereich, so daß solche Verfahren „im Frequenzbereich“ genannt werden.

Zur Verarbeitung im Frequenzbereich wird der gleiche Weg gegangen, der schon in Kap. 4 für Beamformer gegangen wurde: Die Observationssignale x_n^b werden mit der STFT (4.3) transformiert (x_n^f), dort bearbeitet und in den Zeitbereich mit overlap-&-add rücktransformiert. Damit gilt für x_n^f (wobei der obere Index wie schon im letzten Kapitel meist weggelassen wird):

$$y_m^f(t) = \sum_{n=1}^N w_{mn}^f(f) x_n^f = \mathbf{w}_m^{Hf} \cdot \mathbf{x}^f. \quad (5.9)$$

Es wird das Mischungsmodell aus (4.5) $\mathbf{x} = \mathbf{H}\mathbf{s}$ angenommen. Das heißt, es ist identisch mit dem der Beamformer in Kap. 4.1. Von Smaragdis (1998) wurde vorgeschlagen, statt (5.8) in jedem Frequenzabschnitt eine Bestimmungsgleichung zu optimieren.

$$\forall f, i, j \neq i : \mathbb{E}\{g(y_i^f)h(y_j^f)\} = 0 \quad (5.10)$$

Andere Bedingungen aus Kap. 5.1.1 können natürlich auch angewendet werden.

Das Mischungsmodell in jedem Frequenzbereich entspricht damit einer skalaren Mischung. Darin liegt der Hauptvorteil dieser Methode. Abgesehen von einer Erweiterung auf komplexwertige Größen können viele Algorithmen übernommen werden, deren gute Konvergenzeigenschaften für die skalare ICA nachgewiesen wurden. Weiterhin kann die FFT bei der Realisierung der Faltung ausgenutzt werden, was bekanntermaßen den Aufwand von K^2 auf $K \log_2 K$ reduziert.

Die Bedingungen gelten in jedem Frequenzbereich einzeln. Damit kann man wenigstens theoretisch⁷ jede beliebige Entmischungsfiterlänge wählen, indem man die DFT-Länge hinreichend lang wählt. Diese Methode funktioniert also auch in Räumen mit einer realistischen Nachhallzeit.

Die Hauptprobleme der konvolutiven ICA im Frequenzbereich sind das Permutations- und das Skalierungsproblem. Die statistische Unabhängigkeit garantiert nur die Überlagerungsfreiheit, d.h. für

$$y_n = \mathbf{w}_n^T \mathbf{H}\mathbf{s} = \mathbf{q}_n^T \mathbf{s}$$

gilt, daß nur ein Element (z.B. l) von \mathbf{q}_n ungleich null sein darf ($\forall l \neq m : q_{nl} = 0$), wenn y_n von allen anderen Kanälen unabhängig ist. Damit ist $y_n = q_{nm} s_m$, über die

⁷Die obere Grenze, bis zu der dies praktisch möglich ist, wurde von Araki *et al.* (2003) untersucht. Sie ist von der Qualität abhängig, mit der die statistischen Parameter geschätzt werden können. Bei einer endlichen Aufnahme stehen für jeden Frequenzabschnitt weniger Werte zur Verfügung, um die Statistik zu schätzen, wenn die Anzahl der Frequenzabschnitte steigt. Damit verschlechtert sich die Performance oberhalb einer bestimmten Grenze.

Größe von q_{nm} wird keine Aussage gemacht. Das ist als Skalierungsproblem bekannt, weil die Amplitude der Signale nicht rekonstruiert werden kann.

Betrachtet man einen ICA-Algorithmus im Frequenzbereich, gilt dies für jeden Frequenzabschnitt einzeln. Für einen Frequenzabschnitt f_1 ist das Ausgangssignal $y_n^{f_1}$ das um $q_{nm}^{f_1}$ verstärkte Quellsignal $s_m^{f_1}$. In einem anderen Frequenzbereich f_2 kann $q_{nm}^{f_2}$ einen anderen Wert annehmen, was zu einer frequenzabhängigen Gewichtung, also einer Verzerrung führt⁸. Das Skalierungsproblem der skalaren ICA wird jetzt zu einem „Faltungsproblem“. Sind zwei Signale unabhängig, sind auch alle gefalteten Versionen unabhängig!

Weiterhin sagt die Überlagerungsfreiheit nichts darüber aus, welches q_{nm} ungleich null ist. Man weiß also nicht, welche Quelle rekonstruiert wurde. Für eine skalare ICA ist das unbedeutend, weil nur die Reihenfolge der Quellen verloren geht (Permutation), welche ohnehin willkürlich ist. Bei der Schätzung im Frequenzbereich können jedoch in zwei Frequenzabschnitten $q_{nm_1}^{f_1}$ und $q_{nm_2}^{f_2}$ ungleich null sein. Setzt man dann das breitbandige Rekonstruktionssignal y_n^b zusammen, besteht es u.a. aus $q_{nm_1}^{f_1} s_{m_1}^{f_1}$ und $q_{nm_2}^{f_2} s_{m_2}^{f_2}$ und bleibt eine Mischung dieser beiden Quellen. Dieses Problem wird Permutationsproblem genannt.

Diese beiden Probleme stellen derzeit den größten Forschungsbedarf dar und sind noch nicht für den (allgemeinen) Fall gelöst, wenn mehr als zwei Quellen gemischt wurden und getrennt werden sollen.

5.1.4 Hauptachsentransformation

Der erste Schritt der meisten vorgestellten Verfahren ist die Dekorrelation nach (5.4). Sie ist u.a. unter den Namen Karhunen-Loeve-Transformation, Hauptachsentransformation (engl. principle component analysis, PCA), Hotelling-Transformation, Faktoranalyse, Hauptkomponentenanalyse und weiteren bekannt und ist das klassische Werkzeug der statistischen Datenanalyse und Signalverarbeitung. Erstmals wurde sie (bereits) 1901 von Parson (1901) verwendet. Ihr Zweck ist es, für eine multivariate statistische Verteilung eine kleinere Anzahl Variablen mit einer geringeren Redundanz zu finden. Die Redundanz wird dazu im Gegensatz zur ICA nicht durch die statistische Abhängigkeit der Variablen (5.1) sondern über die Kreuzkovarianzen (5.4) gemessen.

Für einen N -dimensionalen (mittelwertfreien) Zufallsvektor werden die Auto- und Kreuzkovarianzen in der Kovarianzmatrix $\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\}$ zusammengefaßt. Sind alle Komponenten eines Zufallsvektors nach (5.4) untereinander unkorreliert, ist die Kovarianzmatrix eine Diagonalmatrix. Damit muß \mathbf{x} durch eine geeignete Transformation $\mathbf{z} = \mathbf{V}\mathbf{x}$ so transformiert werden, daß \mathbf{R}_z eine Diagonalmatrix ist. Darüber hinaus wird oft die Forderungen erhoben, daß die Autokovarianzen der Komponenten von \mathbf{z} gleich sind, also \mathbf{R}_z eine Einheitsmatrix ist.

⁸Dieses Problem ist nicht auf die Verarbeitung im Frequenzbereich beschränkt, weil die Faltung eines Signals nichts an der Gültigkeit von (5.8) ändert. Die FIR-Modelle führen z. B. zu einer Verflachung des Leistungsdichtespektrums, d.h. auch sie rekonstruieren nur eine gefaltete Version der Signale.

Um das zu erreichen, muß das neue Koordinatensystem in die Hauptachsen des Konzentrationsellipsoids der zu der Kovarianzmatrix gehörenden Normalverteilung gelegt werden. Die Hauptachsen sind die Eigenvektoren der Kovarianzmatrix, so daß eine Eigenwertzerlegung von \mathbf{R}_x ausgeführt werden muß. Sie lautet

$$\begin{aligned}\mathbf{R}_x &= \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^H \\ \Leftrightarrow \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^H\mathbf{R}_x\mathbf{U}\mathbf{\Lambda}^{-\frac{1}{2}} &= \mathbf{I} \\ \Leftrightarrow \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^H \cdot \mathbf{E}\{\mathbf{xx}^H\} \cdot \mathbf{U}\mathbf{\Lambda}^{-\frac{1}{2}} &= \mathbf{E}\{\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^H\mathbf{xx}^H\mathbf{U}\mathbf{\Lambda}^{-\frac{1}{2}}\} \\ &= \mathbf{E}\{\mathbf{zz}^H\} = \mathbf{I}\end{aligned}$$

mit $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_N]^H$, der Matrix der unitären Eigenvektoren von \mathbf{R}_x , und der Diagonalmatrix der Eigenwerte $\mathbf{\Lambda}, = \text{diag}[\lambda_1 \dots \lambda_N]$. Die Matrixwurzel, die allgemein nicht eindeutig ist, kann für Diagonalmatrizen leicht angegeben werden $\mathbf{\Lambda}^{\frac{1}{2}} = \text{diag}[\sqrt{\lambda_1} \dots \sqrt{\lambda_N}]$. Die PCA ergibt sich damit zu

$$\mathbf{z} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^H\mathbf{x} = \mathbf{V}\mathbf{x}. \quad (5.11)$$

Die so transformierte Variable hat eine gleiche Varianz in alle Richtungen, weil die Kovarianzmatrix eine Einheitsmatrix ist. Da das Konzentrationsellipsoid eine (Hyper-)Kugel wird, wird dieser Schritt auch *Sphering* genannt⁹

Eine PCA allein genügt nicht für die Trennung der Signale. Denn nicht nur \mathbf{z} sondern auch jedes $\mathbf{z}' = \mathbf{T}\mathbf{z}$, das durch eine unitäre Transformation aus \mathbf{z} hervorgeht, erfüllt die Bedingung, daß die Kreuzkovarianzen verschwinden, wie man leicht sieht:

$$\mathbf{E}\{\mathbf{z}'\mathbf{z}'^H\} = \mathbf{T} \cdot \mathbf{E}\{\mathbf{zz}^H\} \cdot \mathbf{T} = \mathbf{T}\mathbf{T}^H = \mathbf{I}.$$

Damit ist der korrekte Signalvektor eine unitäre Transformation \mathbf{T} von \mathbf{z} . Sie kann nicht aus der Kovarianzmatrix gewonnen werden¹⁰. Dennoch ist sie als Vorverarbeitung sinnvoll, weil sich einerseits das Problem der ICA auf die Suche nach einer orthogonalen Matrix reduziert und andererseits eine Dimensionalitätsreduktion möglich ist.

Die Orthogonalität ergibt sich auch aus folgender Betrachtung:

$$\begin{aligned}\mathbf{I} &= \mathbf{E}\{\mathbf{V}\mathbf{xx}^H\mathbf{V}\} = \mathbf{V}\mathbf{H}\mathbf{R}_s\mathbf{H}^H\mathbf{V} \\ \Rightarrow \mathbf{V}\mathbf{H}(\mathbf{V}\mathbf{H})^H &= \mathbf{I} \quad \Rightarrow \quad \mathbf{V}\mathbf{H} \text{ ist unitär,}\end{aligned}$$

wenn die Quellensignale die gleiche Varianz ($\mathbf{R}_s = \mathbf{I}$) haben. Dann ist $\mathbf{T} = \mathbf{V}\mathbf{H}$ orthogonal und somit $\mathbf{H}^{-1} = \mathbf{T}^H\mathbf{V}$. Zur Invertierung von \mathbf{H} muß also nur noch eine Orthogonalmatrix gefunden werden.

Die Dimensionalitätsreduktion ist möglich, wenn die Anzahl der Observationen N größer als die der Quellen M ist. Dann hat \mathbf{R}_x nur N von Null verschiedene

⁹Ebenfalls ist der Begriff *Weißes* gebräuchlich. Er ist jedoch problematisch, weil die Verwechslungsgefahr zum zeitlichen Dekorrelieren der Signale besteht. Bei einer zeitlichen Dekorrelation entsteht ein flaches Leistungsdichtespektrum, also ein „weißes“ Signal. Hier hingegen werden nur die Kreuzkovarianzen zwischen den Signalen entfernt. Die zeitlichen Abhängigkeiten in den Signalen bleiben erhalten.

¹⁰Die später verwendeten Graphiken Abb. 6.1 und Abb. 6.2 können als Illustration dienen.

Eigenwerte, wenn das additive, unkorrelierte Rauschen hinreichend klein ist. Die Dimensionalität des Problems wird dann dadurch verringert, daß der Signalvektor in den Unterraum transformiert wird, der von den Eigenvektoren aufgespannt wird, die zu den M größten Eigenwerten gehören. Damit kann einerseits das Rauschen verringert werden (Asano *et al.* (2000)). Andererseits kann aus der Dimensionalität des Signalunterraumes die Anzahl der aktiven Geräusche ermittelt werden. Dies wird für die Klassifikation von Maschinensignalen in einem neuen Algorithmus ausgenutzt, der in Kap. 7.2 vorgestellt wird.

5.2 Nichtlineare Dekorrelation

Die in diesem Kapitel vorgestellten Verfahren nutzen (5.3) zur Darstellung der statistischen Unabhängigkeit aus. Um die Algorithmen zu diskutieren, werden die Grundbegriffe der Informationstheorie kurz angerissen.

Die Basisgröße für die gesamte Informationstheorie ist die Entropie. Sie ist für diskrete Verteilungen über folgende Summe definiert.

$$H(y) = - \sum_i P(y = a_i) \log(P(y = a_i)). \quad (5.12)$$

a_i ist ein bestimmtes diskretes Ereignis aus der Menge der möglichen Ereignisse. Der Logarithmus dualis beruht auf der ursprünglichen Definition für die Kanalkodierung von Claude Shannon: Unter vereinfachenden Annahmen hat der minimale binäre Code die Entropie als untere Grenze, die beliebig genau erreicht werden kann, vgl. Cover und Thomas (1991)¹¹.

In der Wahrscheinlichkeitsrechnung gibt die Entropie Auskunft über die Zufälligkeit der Verteilung. Je größer die Entropie ist, desto weniger vorhersagbar und strukturiert ist die Variable. Da dazu die Beziehung zur Kodierungslänge von untergeordneter Bedeutung ist und die Normalverteilung auf Exponentialfunktionen beruht, wird in dieser Arbeit der Logarithmus naturalis in (5.12) verwendet.

Für kontinuierliche Verteilungen muß die Summe in (5.12) in ein Integral übergehen. Daraus kann man die differentielle Entropie definieren¹².

$$H(y) = - \int p(y) \log(p(y)) dy = E\{\log \mathbf{p}_y(\mathbf{y})\} \quad (5.13)$$

Wie unten gezeigt, besitzt die Normalverteilung für eine vorgegebene Varianz die maximale Entropie, analog läßt sich zeigen, daß für eine begrenzte Amplitude die Gleichverteilung die Entropie maximiert.

¹¹Daraus ergibt sich, daß die Kodierungslänge der globale Ansatz für die ICA ist: Mit einer Minimierung der Kodierungslänge für immer auf eine Entropiemaximierung. Aus der Minimierung der Kodierungslänge lassen sich auch die anderen Algorithmen herleiten, die sich nicht aus der statistischen Unabhängigkeit i.e.S. ableiten lassen.

¹²Auf die formalen Probleme der Ableitung der differentiellen Entropie soll an dieser Stelle nicht eingegangen werden, siehe dazu Lee (1997) oder Köhler (2000).

Um (5.1) direkt anzuwenden, wird ein Maß benötigt, daß die Ähnlichkeit zweier Wahrscheinlichkeitsdichten beschreibt. Dazu wird die Kullback-Leibler-Divergenz (relative Entropie) benutzt, die den Abstand zwischen zwei Wahrscheinlichkeitsdichtefunktionen beschreibt. Sie ist für zwei N -dimensionale Verteilungen $p(\mathbf{y})$ und $q(\mathbf{y})$ über folgendes Integral bestimmt

$$D(p||q) = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{q(\mathbf{y})} d\mathbf{y}. \quad (5.14)$$

Die so definierte Kullback-Leibler-Divergenz (KL-Divergenz) kann als Abstand der Wahrscheinlichkeitsdichtefunktionen angesehen werden, weil sie stets nicht-negativ ist und den Wert Null nur für gleiche Verteilungen annimmt.

Die Kullback-Leibler-Divergenz $J(s) = D(p_s||p_{\text{Gauss}})$ zwischen einer Verteilung und der Normalverteilung mit gleicher Varianz heißt Negentropie. Ihr Name ergibt sich daraus, daß sie als relative Entropie zwischen den beiden Verteilungen angesehen werden kann.

$$\begin{aligned} J(\mathbf{y}) = D(p_y||p_{\text{Gauss}}) &= \int p_y(\mathbf{y}) \log \frac{p_y(\mathbf{y})}{p_{\text{Gauss}}(\mathbf{y})} d\mathbf{y} \\ &= \int p_y(\mathbf{y}) \log p_y(\mathbf{y}) d\mathbf{y} - \int p_y(\mathbf{y}) \log p_{\text{Gauss}}(\mathbf{y}) d\mathbf{y} \\ &= H_{\text{Gauss}}(\mathbf{y}) - H(\mathbf{y}) \end{aligned}$$

Dabei ist das Integral $-\int p_y(\mathbf{y}) \log p_{\text{Gauss}}(\mathbf{y}) d\mathbf{y} = H_{\text{Gauss}}(\mathbf{y})$ nach Cover und Thomas (1991) gleich der Entropie einer Normalverteilung mit gleicher Varianz. Da $J(\mathbf{y})$ nicht-negativ ist, gilt $0 \leq H_{\text{Gauss}}(\mathbf{y}) - H(\mathbf{y})$. Damit hat die Normalverteilung die größte Entropie für eine feste Varianz. Ebenfalls läßt sich zeigen, daß die multivariate Negentropie invariant gegenüber orthogonalen Transformationen des Zufallsvektors ist: $J(\mathbf{T}\mathbf{x}) = J(\mathbf{x})$ mit $\mathbf{T}\mathbf{T}^T = \mathbf{I}$. Weiterhin ändert die Varianz die Negentropie nicht, sie ist also skalierungsinvariant.

Nach (5.2) sind für einen Zufallsvektor mit statistisch unabhängigen Komponenten die Verbundwahrscheinlichkeitsdichte und das Produkt der Randwahrscheinlichkeitsdichten gleich. Bei statistisch abhängigen Variablen unterscheiden sie sich, wobei der Unterschied mit der Kullback-Leibler-Divergenz (5.14) quantifiziert werden kann (gemeinsame Information):

$$I(y_1, \dots, y_N) = D(p_{\mathbf{y}}||\prod p_{y_n}) = \int p_{\mathbf{y}}(\mathbf{y}) \log \frac{p_{\mathbf{y}}(\mathbf{y})}{\prod p_{y_n}(y_n)} d\mathbf{y}. \quad (5.15)$$

Leicht ersichtlich folgt daraus (5.16) für die Entropie.

$$I(y_1, \dots, y_N) = \sum_{n=1}^N H(y_n) - H(\mathbf{y}). \quad (5.16)$$

Da es das Ziel vieler Verfahren der blinden Quellentrennung ist, die statistische Abhängigkeit zu minimieren, steht über die gemeinsame Information ein effizientes Hilfsmittel für deren Quantifizierung zur Verfügung.

Die gemeinsame Information läßt sich offensichtlich minimieren, indem man den ersten Term, die Entropie jedes Kanals, minimiert oder den zweiten Term, die Entropie der multivariaten Verteilung, maximiert. Die erste Methode bedeutet die Maximierung der Negentropie oder der „Nichtgaußförmigkeit“ (nongaussianity) der Randverteilungen (Hyvärinen und A. (1999)). Diese Verfahren sind unter dem Namen FastICA bekannt, weil ein schneller Algorithmus durch die Ausnutzung des Newton-Verfahrens entwickelt wurde.

Die zweite Methode heißt Maximum-Likelihood-ICA oder Infomax, weil die Maximierung der Entropie äquivalent mit der Maximierung des Informationsflusses durch einen nichtlinearen Kanal ist (Bell und Sejnowski (1995), Amari *et al.* (1996)). Wie in mehreren Arbeiten nachgewiesen wurde, sind für geweißte Signale alle derartigen Verfahren äquivalent und kommen darum zum gleichen Ergebnis (Lee (1998), Haykin (2000)).

5.2.1 Maximierung der Negentropie (FastICA)

Die Signale \mathbf{z} seien mit einer PCA, Kap. 5.1.4 geweißt (d.h. $E\{\mathbf{z}\mathbf{z}^H\} = \mathbf{I}$), so daß eine Transformation mit einer orthogonalen Matrix \mathbf{T} genügt, um die statistische Unabhängigkeit zu erreichen. Da sich die Entropie und die Negentropie nur um eine Konstante unterscheiden, kann man (5.16) auch mit der Negentropie (für $\mathbf{y} = \mathbf{T}\mathbf{z}$ bzw. $y_n = \mathbf{t}_n^H \mathbf{z}$ mit \mathbf{t}_n^H der n -ten Zeile der Rotationsmatrix \mathbf{T}) schreiben:

$$I(y_1, \dots, y_N) = - \sum_{n=1}^N J(\mathbf{t}_n^H \mathbf{z}) - J(\mathbf{T}\mathbf{z}) = - \sum_{n=1}^N J(\mathbf{t}_n^H \mathbf{z}) + \text{const.} \quad (5.17)$$

Das letzte Gleichheitszeichen gilt aufgrund der Invarianz der Negentropie gegenüber einer orthogonalen Transformation. Die Negentropie ist definitionsgemäß nicht-negativ. Damit muß die Negentropie jedes einzelnen Ausgangssignals maximal sein, um die gemeinsame Information zu minimieren, so daß die Ausgangssignale auch geschätzt werden können (Deflationsalgorithmen). Dies ist in den Anwendungsgebieten dieser Arbeit nützlich, weil damit besonders interessante Quellensignale, z.B. das Nutzsignal, einzeln geschätzt werden können. Im folgenden soll darum auch nur ein Signal geschätzt werden. Die Optimierungsfunktion vereinfacht sich zu $J(\mathbf{t}^H \mathbf{z})$.

Die Wirkungsweise kann man auch aus dem zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung herleiten. Er besagt, daß eine Überlagerung unendlich vieler unabhängiger, identisch verteilter Vektoren \mathbf{z}_i mit gleicher Kovarianzmatrix \mathbf{C}_z und gleichem Erwartungswert \mathbf{m}_z gegen eine multivariate Normalverteilung strebt (Schwarz (2002)).

Seien s_1 und s_2 statistisch unabhängige Zufallsvariablen mit gleicher, nicht gaußförmiger Verteilung mit der Negentropie $J(s_m) = D(p_{s_m} || p_{Gauss})$. Bildet man nun eine beliebige Linearkombination dieser Variablen $y = \mathbf{q}^T [s_1 s_2]^T$, folgt aus dem zentralen Grenzwertsatz, daß deren Negentropie geringer ist.

$$D(p_y || p_{Gauss}) < D(p_{s_m} || p_{Gauss}) \quad (5.18)$$

Die Verteilung einer Linearkombination von statistisch unabhängigen Zufallsvariablen ähnelt stärker einer Normalverteilung als die Ausgangsverteilungen. Da die Negentropie diese Abweichung beschreibt, wird sie genau dann maximal, wenn keine lineare Überlagerung der statistischen unabhängigen Quellen vorliegt (5.18). Oder anders ausgedrückt: Für jeden Vektor \mathbf{t} , der die Negentropie maximiert, gilt $\mathbf{q}_n = \mathbf{t}^H \mathbf{H} = [0 \dots q_{mn} \dots 0]^T$ und das Ausgangssignal ist proportional dem Quellsignal s_m : $y_n = q_{mn} s_m$.

Beschreibung der Negentropie

Formal exakt kann die Negentropie nur dann berechnet werden, wenn die Wahrscheinlichkeitsdichtefunktion bekannt ist. Eine Approximation ist aber auch aus den wesentlich leichter zu schätzenden Kumulanten möglich. Dazu wird die Wahrscheinlichkeitsdichtefunktion durch eine Reihe entwickelt (z.B. Gram-Charlier-Entwicklung, Edgeworth-Entwicklung).

Als Basisfunktion der Gram-Charlier-Reihe dient eine Normalverteilung

$$\phi = \frac{e^{-y^2/2}}{\sqrt{2\pi}}.$$

Die Ableitungen sind

$$\frac{\partial^l \phi(y)}{\partial y^l} = (-1)^l H_l(y) \phi(y).$$

mit den Chebyshev-Hermite-Polynomen H , welche die angenehme Eigenschaft der Orthogonalität besitzen:

$$\int \phi(y) H_k(y) H_l(y) dy = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$$

Damit ergibt sich folgende (der Taylorreihe ähnliche) Reihe bis zum vierten Glied

$$p_y(y) \approx \phi(y) \left(1 + \kappa_3 \frac{H_3(y)}{3!} + \kappa_4 \frac{H_4(y)}{4!} \right).$$

Der nicht-gaußförmige Anteil der Wahrscheinlichkeitsdichtefunktion wird durch Kumulanten höherer Ordnung beschrieben (Koeffizienten der Taylor-Reihenentwicklung der kumulantengenerierenden Funktion, siehe Nikias und Petropulu (1993)). Wird die Reihe nach dem vierten Glied abgebrochen, werden nur die Kurtosis und Schiefe benutzt, die im reellen (mittelwertfreien) Fall wie folgt definiert sind.

$$\kappa_3(y) = E\{y^3\} \quad \text{Schiefe} \quad (5.19)$$

$$\kappa_4(y) = \text{kurt}(y) = E\{y^4\} - 3 \quad \text{Kurtosis} \quad (5.20)$$

Für die Entropie und die Negentropie ergeben sich somit die folgende Kumulantendarstellungen:

$$H(y) = - \int \phi(\xi) \log \phi(\xi) d\xi + \frac{\kappa_3^2}{2 \cdot 3!} + \frac{\kappa_4^2}{2 \cdot 4!} \quad (5.21)$$

$$J(y) = \frac{\kappa_3^2}{2 \cdot 3!} + \frac{\kappa_4^2}{2 \cdot 4!} = \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} \kappa_4^2. \quad (5.22)$$

Verallgemeinert man dieses Konzept, kann man die Wahrscheinlichkeitsdichtefunktion und damit die Negentropie auch durch beliebige nichtlineare Funktionen beschreiben (vgl. Hyvärinen *et al.* (2001)), wenn man anstatt der Chebishev-Hermite-Polynome andere orthogonale Funktionen verwendet, für die gilt:

$$\int \phi(y) F^k(y) F^l(y) dy = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$$

Dann wird die Wahrscheinlichkeitsdichtefunktion analog approximiert:

$$p_y(y) \approx \phi(y) \left(1 + \sum c_i F^i(y) \right).$$

Ein solches Funktionensystem kann man sich aus beliebigen linear unabhängigen Funktionen g_i konstruieren, indem man sie geeignet orthogonalisiert. Das einfachste System erhält man, wenn das System aus einer ungeraden Funktion g_1 und einer geraden Funktion g_2 aufgebaut ist. g_1 beschreibt den schiefsymmetrischen Anteil der Wahrscheinlichkeitsdichtefunktion und g_2 den symmetrischen nichtgaußförmigen Anteil. Nach der Orthogonalisierung erhält man dann unter Verwendung der Definition der Negentropie

$$J(y) \approx (k_1 E\{g_1(y)\})^2 + (k_2 (E\{g_2(y)\} - E\{g_2(y_{\text{Gauss}})\}))^2.$$

Für die meisten Fälle der Signalverarbeitung kann von einer symmetrischen Verteilung ausgegangen werden, so daß g_1 nicht betrachtet zu werden braucht.

$$J(y) \approx (k_2 (E\{g_2(y)\} - E\{g_2(y_{\text{Gauss}})\}))^2 \quad (5.23)$$

Gegenüber einer kumulantenbasierten Approximation hat dieser Weg den Vorteil, daß man durch eine geeignete Wahl der Funktion g_2 die Robustheit des Schätzers einstellen kann. Hat g_2 einen flacheren Anstieg als y^4 fallen starke Ausreißer weniger stark ins Gewicht, was den Schätzer robuster macht. Typische Funktionen, die dies erfüllen, sind in Hyvärinen *et al.* (2001) gegeben (z.B. für leptokurtische (Supergaußsche) Verteilungen $g_2^+ = a_1 / \cosh y$ oder für platykurtische (Subgaußsche) Verteilungen $g_2^- = a_1 \cosh y \cdot e^{y^2/2}$). Die Verteilungen, die zur Modellierung platykurtischer Verteilungen verwendet werden, werden auch in Kap. 7.1 diskutiert. Eine einfache Threshold-Funktion wurde von Mathis *et al.* (2000) vorgeschlagen.

Weiterhin kann man die Kullback-Leibler-Divergenz und damit die auch Negentropie auch über Quantile schätzen. Ein solcher Algorithmus wurde von Pham (2000) vorgeschlagen.

Minimierung der Negentropie

Da die in Hyvärinen *et al.* (2001) angegebenen Funktionen von reellwertigen Argumenten ausgehen, hier jedoch komplexe Zufallsvektoren vorliegen, muß man das Konzept erweitern, indem man g_2 auf das Betragsquadrat von $\mathbf{t}^H \mathbf{z}$ anwendet (Bingham und Hyvärinen (2000)).

Um (5.17) zu erfüllen, werden die Minima von J gesucht. Am einfachsten kann man sie mit einem Gradientenverfahren bestimmen, wobei wieder der komplexe Gradient von Brandwood (1983) genutzt wird, siehe Kap. 4.3.1.

$$\nabla_{\mathbf{t}^*} J = 2 \left[\mathbb{E}\{g_2(|\mathbf{t}^H \mathbf{z}|^2)\} - \mathbb{E}\{g_2(y_{\text{Gauss}})\} \right] \mathbb{E}\{g_2'(|\mathbf{t}^H \mathbf{z}|^2) \cdot \mathbf{z} \cdot (\mathbf{t}^H \mathbf{z})^*\}$$

Die Größe des ersten Faktors hat keinen Einfluß auf die Lage der stationären Punkte, lediglich sein Vorzeichen verändert die Stabilität. Es entspricht dem Vorzeichen der Kurtosis und kann damit leicht a priori angegeben werden, wenn Vorwissen darüber vorliegt, ob die Quellensignale lepto- oder platykurtisch sind. Für den stochastischen Gradienten ist eine Normierung von \mathbf{t} auf die Länge 1 nötig, weil das globale Maximum von J auf dem Koordinatenursprung liegt. Diese Nebenbedingung ist angemessen, weil in Kap. 5.1.4 gezeigt wurde, daß die Rotationsmatrix eine unitäre Matrix sein muß.

Damit stellt sich die Frage, ob das Optimierungsproblem auch mit der Lagrange-Methode algebraisch gelöst werden kann. Die Nebenbedingung lautet $h_1 : \mathbf{t}^H \mathbf{t} = 1$, ihr Gradient ist $\nabla_{\mathbf{t}^*} h_1 = \mathbf{t}$. Damit muß ein Kuhn-Tucker-Punkt (Definition, siehe Kap. 4.3.1) die Bedingung erfüllen:

$$\mathbb{E}\{g_2'(|\mathbf{t}^H \mathbf{z}|^2) \cdot \mathbf{z} \cdot (\mathbf{t}^H \mathbf{z})^*\} = \alpha \mathbf{t}. \quad (5.24)$$

Durch die Nicht-Linearität von g_2 läßt sich keine algebraische Lösung finden. Gleichung (5.24) ermöglicht jedoch eine iterative Lösung, was einem modifizierten Newton-Verfahren entspricht, so daß dessen gute Konvergenzeigenschaften ausgenutzt werden. Weiterhin zeigt (5.24), daß ein optimaler Entmischungsvektor in die selbe Richtung wie der Gradient der Kostenfunktion an dieser Stelle zeigen muß.

Die iterative Lösung nach dem Newton-Verfahren erfordert die Ableitung von (5.24), d.h. die Jacobi-Matrix \mathbf{J} von $\mathbf{t}^H \mathbf{z}$:

$$\begin{aligned} \nabla_{\mathbf{t}} (\mathbb{E}\{g_2'(|\mathbf{t}^H \mathbf{z}|^2) \mathbf{z} \mathbf{z}^H \mathbf{t}\} - \alpha \mathbf{t}) &= \mathbb{E}\{g_2''(|\mathbf{t}^H \mathbf{z}|^2) (\mathbf{z} \mathbf{z}^H \mathbf{t}) (\mathbf{t}^H \mathbf{z} \mathbf{z}^H) + \\ &\quad g_2'(|\mathbf{t}^H \mathbf{z}|^2) (\mathbf{z} \mathbf{z}^H)\} + \alpha \mathbf{I} \\ &= \mathbb{E}\{g_2''(|\mathbf{t}^H \mathbf{z}|^2) (\mathbf{z} \mathbf{z}^H) |\mathbf{t}^H \mathbf{z}|^2 + \\ &\quad g_2'(|\mathbf{t}^H \mathbf{z}|^2) (\mathbf{z} \mathbf{z}^H)\} + \alpha \mathbf{I} \\ &= \mathbb{E}\{g_2''(|\mathbf{t}^H \mathbf{z}|^2) |\mathbf{t}^H \mathbf{z}|^2 + \\ &\quad g_2'(|\mathbf{t}^H \mathbf{z}|^2)\} \cdot \mathbf{I} + \alpha \mathbf{I}. \end{aligned}$$

Damit ist die Jacobi-Matrix eine Einheitsmatrix und kann leicht (ohne die Rechenkosten einer „normalen“ Matrixinversion) invertiert werden. Für das Newtonverfahren ergibt sich

$$\tilde{\mathbf{t}}_{n+1} = \mathbf{t}_n - \frac{\mathbb{E}\{g_2'(|\mathbf{t}^H \mathbf{z}|^2) \cdot \mathbf{z} \cdot (\mathbf{t}^H \mathbf{z})^*\} - \alpha \mathbf{t}}{\mathbb{E}\{g_2'(|\mathbf{t}^H \mathbf{z}|^2)\} + \alpha}. \quad (5.25)$$

Gleichung (5.25) garantiert nicht die Einhaltung der Randbedingung, sie garantiert nur, daß der Vektor \mathbf{t} in die Richtung des Gradienten von J zeigt. Darum wird nach jedem Schritt mit (5.26) normiert. Abbildung 5.2 zeigt die Konvergenz des Algorithmus.

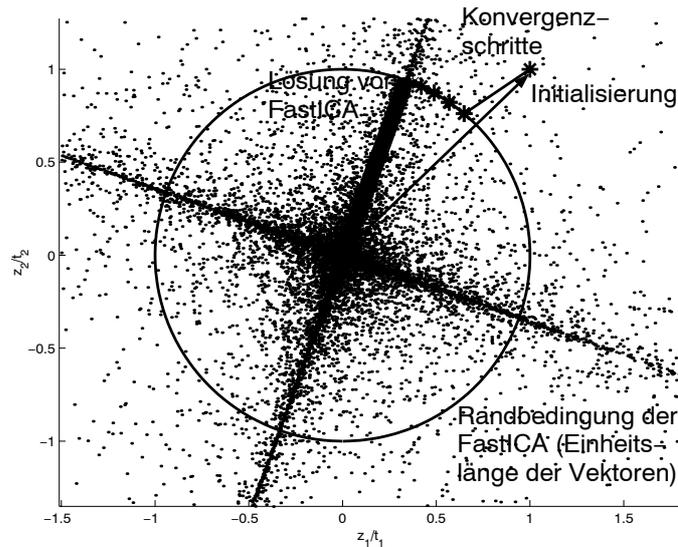


Abb. 5.2: Konvergenz eines FastICA-Algorithmus auf dem Einheitskreis

$$\mathbf{t}_{n+1} = \frac{\tilde{\mathbf{t}}_{n+1}}{|\tilde{\mathbf{t}}_{n+1}|} \quad (5.26)$$

Möchte man mehr als ein Ausgangssignal schätzen, schätzt man ein Signal nach dem anderen und orthogonalisiert jeweils die Signale nach jeder Schätzung (z.B. nach dem Verfahren von Gram-Schmidt), um eine Konvergenz auf den gleichen Kanal zu vermeiden. Dieses Vorgehen wird Deflation genannt. Alternativ dazu kann man auch alle Kanäle gleichzeitig schätzen, was hier jedoch von untergeordneter Bedeutung ist, weil primär nur das Nutzsignal (Maschinensignal) interessiert.

Die FastICA wurde von Hyvärinen und A. (1999) für leptokurtische Verteilungen eingeführt und später von den gleichen Autoren auf beliebige Verteilungen erweitert. Eine Erweiterung auf komplexe Größen erfolgte durch Bingham und Hyvärinen (2000). Für konvolutive Mischungen im Frequenzbereich wurde sie u.a. von Mitia-noudis und Davies (2001) und von Knaak *et al.* (2002) implementiert.

5.2.2 Maximierung der Transinformation (Infomax)

Grundidee der Informationsmaximierung

Für die Ableitung des Algorithmus soll zunächst die Entropie einer Transformation untersucht werden. Im Gegensatz zu den Herleitungen in den Lehrbüchern der ICA muß die Herleitung hier mit komplexwertigen Größen erfolgen, was einige Veränderungen zur Folge hat.

Sei $\mathbf{y} = \mathbf{g}(\mathbf{z})$ eine vektoriell- und komplexwertige Funktion eines Vektors. Für die Wahrscheinlichkeitsdichten gilt nun nach MacKay (2001)

$$p_{\mathbf{y}} = \frac{p_{\mathbf{z}}}{|\det \mathbf{J}_{\mathbf{g}}|}$$

mit \mathbf{J} , der Jacobi-Matrix von \mathbf{g} . Für die Entropie ergibt sich

$$H(\mathbf{y}) = H(\mathbf{z}) + \log |\det \mathbf{Jg}|.$$

Für den linearen Fall $\mathbf{g} = \mathbf{Tz}$ gilt nach Therrien (1992):

$$p_{\mathbf{y}} = \frac{p_{\mathbf{z}}}{\det \mathbf{T}(\det \mathbf{T})^*} \quad H(\mathbf{y}) = H(\mathbf{z}) + \log(\det \mathbf{T}(\det \mathbf{T})^*).$$

Damit kann man die gemeinsame Information in (5.16) vereinfachen, wenn man wieder annimmt, daß die Signale zunächst durch eine PCA geweißt wurden¹³, so daß in diesem Schritt die Multiplikation mit einer orthogonalen Matrix \mathbf{T} ausreicht.

$$\begin{aligned} I(y_1, \dots, y_N) &= \sum H(y_n) - H(\mathbf{y}) \\ &= \sum H(y_n) - H(\mathbf{z}) - \log(\det \mathbf{T}(\det \mathbf{T})^*) \\ &= \sum H(y_n) - \log(\det \mathbf{T}(\det \mathbf{T})^*) + \text{const} \\ &= \sum \mathbb{E}\{\log p_n(\mathbf{t}_n^T \mathbf{z})\} - \log(\det \mathbf{T}(\det \mathbf{T})^*) + \text{const} \end{aligned}$$

Die Verwendung von $\mathbf{t}_n^T \mathbf{z}$ für das Skalarprodukt hat formale Gründe, weil \mathbf{t}_n^T die n -te Zeile der Rotationsmatrix bezeichnet und eine komplexe Konjugation nicht nötig ist.

Die Berechnung des Gradienten ist nur bei Kenntnis der ursprünglichen Wahrscheinlichkeitsdichte möglich. Um ihn weiter zu analysieren, werden zunächst nur reellwertige Zufallsvariablen betrachtet. Dann läßt sich die Ableitung der gemeinsamen Information nach \mathbf{T} leicht beschreiben:

$$\begin{aligned} \nabla_{\mathbf{T}} I &= \nabla_{\mathbf{T}} \sum \mathbb{E}\{\log p_n(\mathbf{t}_n^T \mathbf{z})\} + \nabla_{\mathbf{T}} \log |\det \mathbf{T}| \\ &= \sum \mathbb{E}\{(\log p_n(\mathbf{t}_n^T \mathbf{z}))' \mathbf{z}^T\} - \mathbf{T}^{-T}. \end{aligned}$$

Dabei wird ausgenutzt, daß gilt $\nabla_{\mathbf{T}} \log |\det \mathbf{T}| = \mathbf{T}^{-1T} = \mathbf{T}^{-T}$, was sich durch einfache Umformungen zeigen läßt.

Infomax für reellwertige Signale

Für den reellwertigen Infomax-Algorithmus werden nichtlineare Funktionen g_n , die nach (5.27) definiert sind, und eine vektoruell-wertige Funktion $\mathbf{g} = [g_1(y_1), \dots, g_N(y_N)]^T$ eingeführt. Diese Funktionen werden Aktivierungsfunktionen genannt.

$$g_n = (\log p_n)' = \frac{p_n'}{p_n} \quad (5.27)$$

¹³Das räumliche Weißen ist für die Ableitung des Algorithmus nicht unbedingt nötig. Mehrere Autoren, z.B. Lee (1997), haben eine bessere Konvergenz für die skalare BSS beobachtet, wenn die Signale geweißt sind. Für die konvolutive BSS im Frequenzbereich ist das räumliche Weißen jedoch problematisch, weil es zu einer Verflachung des Leistungsdichtespektrums führt, da die Varianzen eines schmalbandigen Signals x^f der Leistung von x^b in diesem Frequenzbereichs entspricht. Wird das Leistungsdichtespektrum auf einen konstanten Wert gesetzt, entspricht das auch einem zeitlichen Weißen. Darum wird es oft unterlassen, z.B. Sawada *et al.* (2002a).

Die Nichtlinearität müßte (eigentlich) in Abhängigkeit der Wahrscheinlichkeitsdichtefunktion der Quellensignale bestimmt werden. Darum wurden häufige semiparametrische Methoden vorgeschlagen (z.B. von ichi Amari (1997) oder Pearlmutter und Parra (1996)). Wie jedoch von zahlreichen Autoren nachgewiesen wurde (vgl. Hyvärinen *et al.* (2001) für Referenzen), sind die Schätzer auch dann konsistent, wenn die Abweichungen von der realen Wahrscheinlichkeitsdichte hinreichend klein sind. Hinreichend klein sind sie bereits, wenn die vermutete Wahrscheinlichkeitsdichte das richtige Vorzeichen der Kurtosis besitzen¹⁴. Allerdings sind die Konvergenzeigenschaften besser, wenn die Nichtlinearität besser mit den Nichtlinearitäten korrespondiert. Darum wird in Kap. 7.1 die Wahrscheinlichkeitsdichte der Maschinen genauer untersucht und die verwendete Nichtlinearität optimiert.

Im allgemeinen wird für die Wahl von g bei leptokurtischen (supergaußverteilten) Signalen eine Laplace-Verteilung angenommen, die als $p^+ = a/\cosh y$ approximiert wird¹⁵, wodurch sich $g_n^+ = 2 \tanh y_n$ ergibt. (vgl. Abb. 5.3)

Für platykurtische (subgaußverteilte) Signale wird die Verteilung dadurch angenähert, daß man die Normalverteilung mit einem cosh aufweitet und damit breiter macht: $p^- = a \cosh y \cdot e^{y^2/2}$. Damit ergibt sich die Nichtlinearität zu $g_n^- = -\tanh y_n + y_n$. Die optimale Nichtlinearität für Motorsignale wird in Kap. 7.1 hergeleitet.

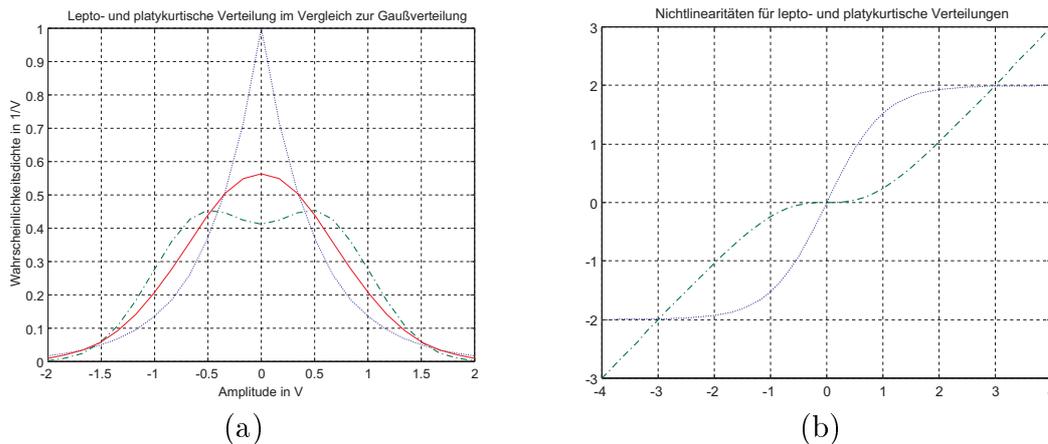


Abb. 5.3: Funktionsverlauf der Verteilungen in (a): p_n^+ (gestrichelt) und p_n^- (gepunktet) sowie der Nichtlinearitäten in (b): g_n^+ (gestrichelt) und g_n^- (gepunktet)

Eine bessere Konvergenz erreicht man nach Amari *et al.* (1996) durch die Verwendung des natürlichen Gradienten¹⁶. Damit ergibt sich für den Gradientenabstieg

¹⁴Dies korrespondiert mit dem Ergebnis des letzten Abschnitts (vgl. Kap. 5.2.1), bei dem ebenfalls eine große Gruppe von Nichtlinearitäten zu konsistenten Schätzern führte. Die Aktivierungsfunktionen entsprechen den Nichtlinearitäten für die Bestimmung der Negentropie.

¹⁵Mit der Annahme $e^{-x} = 1/e^x \approx 1/(e^x + e^{-x}) = 1/\cosh x$.

¹⁶Orthogonale (bzw. hier unitäre) Matrizen bilden im Vektorraum der Matrizen die sog. Grassmann-Mannigfaltigkeit. In ihr gelten die Eigenschaften der euklidischer Geometrie nicht, so daß die Riemann-Geometrie angewendet werden muß. Damit konvergiert, wie von Amari *et al.* (1996) gezeigt wurde, in ihr der Riemann- oder natürliche Gradient deutlich schneller. Der natürliche Gradient ergibt sich allgemein aus dem euklidischen Gradienten über die Metrik des Rie-

(mit reellwertigen Zufallsvektoren):

$$\begin{aligned}\Delta \mathbf{T} &= \alpha \frac{\partial I}{\partial \mathbf{T}} = \alpha (\mathbf{E}\{-\mathbf{g}(\mathbf{T}\mathbf{z})\mathbf{z}^T\} + \mathbf{T}^T)\mathbf{T}^T \mathbf{T} \\ &= \alpha (\mathbf{I} - \mathbf{E}\{\mathbf{g}(\mathbf{T}\mathbf{z})(\mathbf{T}\mathbf{z})^T\})\mathbf{T} \end{aligned} \quad (5.28)$$

$$= \alpha (\mathbf{I} - \mathbf{E}\{\mathbf{g}(\mathbf{y})(\mathbf{y})^T\})\mathbf{T}. \quad (5.29)$$

Diese Gleichung konvergiert gegen stationäre Punkte im Parameterraum, für die gilt:

$$\langle g(y_i) \cdot y_j \rangle = 0 \quad (i \neq j) \quad (5.30)$$

$$\langle g(y_i) \cdot y_j \rangle = 1 \quad (i = j). \quad (5.31)$$

Da Signale geweißt sind, stehen auf der Hauptdiagonalen gleiche Werte; damit minimiert (5.29) die nichtlinearen Korrelationen zwischen den Kanälen. Diese Gleichung stellt also eine direkte Realisierung von (5.3) dar. Eine direkte Implementierung von (5.30) wurde von Chichocki und Unbehauen (1996) vorgeschlagen, indem eine PCA auf $\mathbf{E}\{g(\mathbf{y})\mathbf{y}^T\}$ angewendet wird („Nonlinear PCA“).

Gleichung (5.29) wurde ursprünglich auf zwei anderen Wegen hergeleitet. Zum einen wurde sie von Amari (1997) gefunden, indem er einen Maximum-Likelihood-Schätzer für \mathbf{T} für die Maximierung der Unabhängigkeit beschreibt. Die stationären Punkte von (5.29) sind dann die Maxima der logarithmischen Likelihood-Funktion.

Zum anderen kommt man zu dieser Gleichung über die Theorie künstlicher neuronaler Netze. Ein Neuron ist dann informationsoptimiert, wenn seine Gewichte so eingestellt sind, daß der Informationsfluß von den Eingangs- zu den Ausgangssignalen maximal ist. Die Maximierung der Transinformation zwischen Ein- und Ausgang ist nach Barlow *et al.* (1989) äquivalent mit einer Redundanzreduktion und somit einer Zerlegung in unabhängige Quellen. Wenn \mathbf{g} die Aktivierungsfunktionen von N Neuronen sind, beschreibt (5.29) die Transinformation dieses Netzes. Damit erklärt sich die verwendete Bezeichnung Infomax. Durch diese Überlegung wurde der Algorithmus von Bell und Sejnowski (1995) hergeleitet und popularisiert.

Infomax für komplexe Zufallsvariablen

Die Wahrscheinlichkeitsdichtefunktion einer komplexen Zufallsvariablen, die im folgenden genauer definiert wird, ist als Abbildung $\mathbb{C}^N \rightarrow \mathbb{R}$ keine analytische Funktion. Um dennoch einen Gradienten definieren zu können, wird der komplexe Gradient von Brandwood (1983) benutzt, der schon im vorangegangenen Kapitel (Kap. 4.3.1) Anwendung gefunden hat. Da die gemeinsame Information eine Matrix als Argument hat, muß der konjugiert-komplexe Gradient auf eine komplexe „Ableitungsmatrix“ $\nabla_{\mathbf{T}^*} = [\nabla_{\mathbf{t}_1^*}, \dots, \nabla_{\mathbf{t}_N^*}]$ erweitert werden.

mantraumes durch $\nabla_{\mathbf{T},R} = G_R^{-1} \nabla_{\mathbf{T},E}$. Den Riemanngradienten für Grassmann-Mannigfaltigkeiten erhält man durch die rechtsseitige Multiplikation mit $\mathbf{T}^H \mathbf{T}$. Das entspricht dem relativen Gradienten, der von Cardoso (1998) eingeführt wurde.

Um einen Gradientenabstieg für komplexe Zufallsvariablen zu erklären, muß man die Wahrscheinlichkeitsdichte für komplexe Zufallsvariablen definieren. Dazu wird angenommen, daß die Phase vorwiegend von der Position des Fensters der STFT abhängig und somit etwas willkürlich ist, so daß die Wahrscheinlichkeitsdichtefunktion auf dem Betrag erklärt wird.

Definition 1. *Es sei $Z \in \mathbb{C}$ eine komplexe Zufallsvariable und P die Auftretenswahrscheinlichkeit von Z im Intervall $[x, x + \varepsilon]$. Existiert eine nicht-negative Funktion $p(x)$, für die gilt: $p(x) = \lim_{\varepsilon \rightarrow \infty} \frac{1}{\varepsilon} P(x < |Z| < x + \varepsilon)$, so heißt $p(x)$ Wahrscheinlichkeitsdichtefunktion.*

Die zu minimierende Funktion lautet nun:

$$\begin{aligned} I(y_1, \dots, y_N) &= \sum_{n=1}^N \mathbb{E}\{\log p_n(|\mathbf{t}_n^T \mathbf{z}|)\} - \log(\det \mathbf{T}(\det \mathbf{T})^*) + \text{const} \\ &= I_1 - I_2 + \text{const} \end{aligned}$$

Nun gilt:

Lemma 4. *Es seien*

$$\tilde{g}_n(y_n) = \frac{1}{2} g_n(|y_n|) \cdot e^{j\theta_y} \quad (5.32)$$

die komplexwertigen Aktivierungsfunktionen mit der Phase θ_y . Nun gilt die komplexe Infomax-Regel:

$$\Delta \mathbf{T} = \alpha (\mathbf{I} - \mathbb{E}\{\tilde{\mathbf{g}}(\mathbf{y}) \mathbf{y}^H\}) \mathbf{T}. \quad (5.33)$$

Beweis. Zunächst wird die Ableitung für den zweiten Term I_2 untersucht.

$$\begin{aligned} \nabla_{\mathbf{T}^*} I_2 &= \nabla_{\mathbf{T}^*} \log(\det \mathbf{T}(\det \mathbf{T})^*) \\ &= \frac{1}{\det \mathbf{T}(\det \mathbf{T})^*} \det \mathbf{T} \cdot \nabla_{\mathbf{T}^*} (\det \mathbf{T})^* \\ \nabla_{\mathbf{T}^*} (\det \mathbf{T})^* &= \det \mathbf{T}^* \cdot \mathbf{T}^{*-T} \\ \nabla_{\mathbf{T}^*} I_2 &= \mathbf{T}^{-H}. \end{aligned}$$

Den zweiten Term kann man zunächst als Matrix mit den komplexen Gradienten als Spalten auffassen. Damit ergibt sich für die Summe:

$$\begin{aligned} \nabla_{\mathbf{T}^*} I_1 &= \sum_{n=1}^N \nabla_{\mathbf{T}^*} \mathbb{E}\{\log p_n(|\mathbf{t}_n^T \mathbf{z}|)\} \\ &= [\nabla_{\mathbf{t}_1^*} I_1 \quad \mathbf{0}_{(2)} \dots \mathbf{0}_{(N)}] + [\mathbf{0}_{(1)} \quad \nabla_{\mathbf{t}_2^*} I_1 \quad \mathbf{0}_{(3)} \dots \mathbf{0}_{(N)}] + \dots \\ &\quad + [\mathbf{0}_{(1)} \dots \mathbf{0}_{(N-1)} \quad \nabla_{\mathbf{t}_N^*} I_1] \\ &= [\nabla_{\mathbf{t}_1^*} I_1 \dots \nabla_{\mathbf{t}_N^*} I_1]. \end{aligned}$$

Betrachtet man nun einen komplexen Gradienten ergibt sich:

$$\nabla_{\mathbf{t}_n^*} \mathbb{E}\{\log p_n(|\mathbf{t}_n^T \mathbf{z}|)\} = \mathbb{E}\{\nabla_{|y_n|} \log p_n(|y_n|) \cdot \nabla_{\mathbf{t}_n^*} |y_n|\}.$$

Die äußere Ableitung ist rein reell, so daß man die Definition aus (5.27) benutzen kann:

$$\nabla_{|y_n|} \log p_n(|y_n|) = g_n(|y_n|).$$

Die innere Ableitung ergibt sich wie folgt:

$$\begin{aligned} \nabla_{\mathbf{t}_n^*} |y_n| &= \nabla_{\mathbf{t}_n^*} \sqrt{(\mathbf{t}_n^T \mathbf{z} \cdot (\mathbf{t}_n^T \mathbf{z})^*)} \\ &= \nabla_{\mathbf{t}_n^{T*}} \sqrt{(\mathbf{t}_n \mathbf{z} \cdot \mathbf{z}^{*T} \mathbf{t}_n^*)} \\ &= \nabla_{\mathbf{t}_n^*} \sqrt{(\mathbf{t}_n^T \mathbf{z} \cdot \mathbf{z}^H \mathbf{t}_n^*)} = \frac{1}{2\sqrt{(\mathbf{t}_n^T \mathbf{z} \cdot \mathbf{z}^H \mathbf{t}_n^*)}} \cdot \mathbf{t}_n^T \mathbf{z} \cdot \mathbf{z}^H \\ &= \frac{1}{2|y_n|} \cdot y_n \cdot \mathbf{z}^H = \frac{1}{2} e^{j\theta_y} \cdot \mathbf{z}^H. \end{aligned}$$

Der Faktor $\frac{1}{2}$ kann in die Nichtlinearität gezogen werden. Die Ableitung von I ist mit dem Vektor $\tilde{\mathbf{g}}$ der komplexen Aktivierungsfunktionen \tilde{g}_n aus (5.32)

$$\nabla_{\mathbf{T}^*} I = \nabla_{\mathbf{T}^*} I_1 + \nabla_{\mathbf{T}^*} I_2 = \mathbb{E}\{\tilde{\mathbf{g}}(|y_n|) \cdot \mathbf{z}^H\} - \mathbf{T}^{-H}.$$

Für den Gradientenabstieg ergibt sich mit dem natürlichen Gradienten:

$$\begin{aligned} \Delta \mathbf{T} &= \alpha(-\mathbb{E}\{\tilde{\mathbf{g}}(\mathbf{y}) \cdot \mathbf{z}^H\} + \mathbf{T}^{-H}) \mathbf{T}^H \mathbf{T} \\ &= \alpha(\mathbf{I} - \mathbb{E}\{\tilde{\mathbf{g}}(\mathbf{y}) \mathbf{y}^H\}) \mathbf{T}. \end{aligned}$$

q.e.d.

Damit ist \tilde{g}_n als Aktivierungsfunktion geeignet¹⁷. Nun kann man (5.33) weiter vereinfachen, indem man ein einzelnes Element betrachtet. Man erhält

$$\tilde{g}_n(y) y_n^* = g_n(|y_n|) e^{j\theta_y} \cdot |y_n| \cdot e^{-j\theta_y} = g_n(|y_n|) |y_n|$$

und damit den Infomax in Polarkoordinaten

$$\Delta \mathbf{T} = \alpha(\mathbf{I} - \mathbb{E}\{\mathbf{g}(|\mathbf{y}|) |\mathbf{y}|^H\}) \mathbf{T}. \quad (5.34)$$

Erstmalig wurde die Erweiterung des Infomax-Algorithmus auf komplexe Zahlen von Smaragdís (1998) vorgeschlagen. Seine Aktivierungsfunktion war jedoch „etwas ad-hoc“ (η ist ein Skalierungsfaktor):

$$\tilde{g}_n(y_n) = \tanh[\eta \cdot \Re(y_n)] + j \cdot \tanh[\eta \cdot \Im(y_n)]. \quad (5.35)$$

Zum einen gilt diese Aktivierungsfunktion nur für leptokurtische Zufallsvariablen und ist damit nicht allgemeingültig. Zum anderen müssen die stationären Punkte aus (5.30) auch hier gelten. Betrachtet man die Erwartungswerte auf der Hauptdiagonalen, erhält man:

$$\begin{aligned} \mathbb{E}\{\tilde{g}(y_i) \cdot y_i^*\} &= \mathbb{E}\{[\tanh(\eta \cdot \Re(y_i)) + j \cdot \tanh(\eta \cdot \Im(y_i))] \cdot [\Re(y_i) - j \cdot \Im(y_i)]\} = 1 \\ &= \mathbb{E}\{\tanh(\eta \cdot \Re(y_i)) \cdot \Re(y_i) + j \cdot \tanh(\eta \cdot \Im(y_i)) \cdot \Re(y_i) \\ &\quad + \tanh(\eta \cdot \Im(y_i)) \cdot \Im(y_i) - j \cdot \tanh(\eta \cdot \Re(y_i)) \cdot \Im(y_i)\}. \end{aligned} \quad (5.36)$$

¹⁷Der Faktor 1/2 ändert das grundsätzliche Verhalten nicht und wird im weiteren weggelassen.

Zerlegt man diese Gleichung in Real- und Imaginärteil, so ergeben sich die folgenden Beziehungen:

$$\begin{aligned}\langle \tilde{g}[\Re(y_i)] \cdot \Re(y_i) + \tilde{g}[\Im(y_i)] \cdot \Im(y_i) \rangle &= 1 \\ \langle \tilde{g}[\Im(y_i)] \cdot \Re(y_i) - \tilde{g}[\Re(y_i)] \cdot \Im(y_i) \rangle &= 0.\end{aligned}$$

Aus der zweiten Gleichung ist sofort ersichtlich, daß $\Re(y_i)$ und $\Im(y_i)$ ebenfalls statistisch unabhängig sein sollen. Diese Bedingung ist zu streng, da nicht davon ausgegangen werden kann, daß die Real- und Imaginärteile der Quellensignale jeweils statistisch unabhängig sind.

Murata *et al.* (1999) und Davies (2001) verwenden zur Bestimmung der Entmischungsmatrix eine abgewandelte Form des natürlichen Gradienten:

$$\Delta \mathbf{W} = \mu [\text{diag}(\langle \tilde{g}(\mathbf{y}) \cdot \mathbf{y}^H \rangle) - \langle \tilde{g}(\mathbf{y}) \cdot \mathbf{y}^H \rangle] \cdot \mathbf{W} \quad (5.37)$$

mit der Aktivierungsfunktion von Smaragdis (5.35). Mukai *et al.* (2001) schlugen eine alternative Aktivierungsfunktion vor:

$$\tilde{g}(\mathbf{y}) = \frac{1}{1 + e^{-\Re(\mathbf{y})}} + j \cdot \frac{1}{1 + e^{-\Im(\mathbf{y})}}. \quad (5.38)$$

Bei diesem Algorithmus werden die Werte auf der Hauptdiagonalen nicht in die Optimierung miteinbezogen, so daß das Problem, daß die Real- und Imaginärteile der einzelnen Signale statistisch unabhängig sein müssen, nicht mehr auftritt. Auch diese Aktivierungsfunktion hat den Nachteil, daß sie nur auf leptokurtische Zufallsvariablen anwendbar ist.

5.2.3 Minimierung der Kreuzkumulanten (Jade)

Der Formulierung der statistischen Unabhängigkeit nach Cardoso (1998) oder Mansour und Ohnishi (1999) erfolgt über die Kumulanten. Die Kumulanten vierter Ordnung werden für komplexe Zufallsvariablen wie folgt definiert:

$$\begin{aligned}\text{Cum}(y_i, y_j, y_k, y_l) &= E\{y_i y_j^* y_k^* y_l\} - E\{y_i y_j^*\} E\{y_l y_k^*\} \\ &\quad - E\{y_i y_k^*\} E\{y_l y_j^*\} - E\{y_i y_l\} E\{y_j y_k\}^* \quad (5.39)\end{aligned}$$

Gleichung (5.39) beschreibt einen Tensor, so daß sie von hier an als Kumulantenmatrizen notiert wird

$$\left[\hat{\mathbf{Q}}_{kl}^y \right]_{ij} = \text{Cum}(y_i, y_j, y_k, y_l).$$

für alle $k, l = 1, \dots, m$. Es gibt also maximal m^2 solcher Kumulantenmatrizen.

Aus (5.3) folgt, daß die Kumulanten für statistisch unabhängige Signale bereits dann verschwinden, wenn zwei Indizes sich unterscheiden. Damit ist, wenn \mathbf{s} wie in den vorangegangenen Kapiteln als schmalbandige Signalkomponente definiert ist,

$$\text{Cum}(s_i, s_j, s_k, s_l)|_{i=j=k=l} = \text{kurt}(s_k)$$

und null in allen anderen Fällen. Damit ist es mit den zuvor vorgestellten Verfahren äquivalent, die Kreuzkumulanten für \mathbf{y} zu minimieren.

Für $\mathbf{x} = \mathbf{H}\mathbf{s}$ gilt

$$\text{Cum}(x_i, x_j, x_k, x_l) = \sum_{u=1}^m \text{kurt}(s_u) h_{iu} h_{ju}^* h_{ku}^* h_{lu}, \quad (5.40)$$

wobei $\text{kurt}(s_u)$ die Kurtosis der Quelle s_u darstellt; $\hat{\mathbf{Q}}_{kl}^x$ ergibt sich also zu

$$\begin{aligned} \hat{\mathbf{Q}}_{kl}^x &= \begin{bmatrix} \sum_{u=1}^m \text{kurt}(s_u) h_{1u} h_{1u}^* h_{ku}^* h_{lu} & \sum_{u=1}^m \text{kurt}(s_u) h_{1u} h_{2u}^* h_{ku}^* h_{lu} & \dots \\ \sum_{u=1}^m \text{kurt}(s_u) h_{2u} h_{1u}^* h_{ku}^* h_{lu} & & \\ \vdots & & \ddots \end{bmatrix} \\ &= \sum_{u=1}^m \text{kurt}(s_u) h_{ku}^* h_{lu} \begin{bmatrix} h_{1u} h_{1u}^* & h_{1u} h_{2u}^* & \dots \\ h_{2u} h_{1u}^* & & \\ \vdots & & \ddots \end{bmatrix}. \end{aligned} \quad (5.41)$$

Aus (5.41) läßt sich erkennen, daß die $\hat{\mathbf{Q}}_{kl}^x$ hermitisch sind. Weiteres Umformen ergibt:

$$\begin{aligned} \hat{\mathbf{Q}}_{kl}^x &= \sum_{u=1}^m \begin{bmatrix} h_{1u} \\ \vdots \\ h_{mu} \end{bmatrix} \text{kurt}(s_u) h_{ku}^* h_{lu} \begin{bmatrix} h_{1u} \\ \vdots \\ h_{mu} \end{bmatrix}^H \\ &= \begin{bmatrix} h_{11} & \dots & h_{1m} \\ \vdots & \ddots & \vdots \\ h_{m1} & \dots & h_{mm} \end{bmatrix} \cdot \begin{bmatrix} \text{kurt}(s_1) h_{k1}^* h_{l1} & & 0 \\ & \ddots & \\ 0 & & \text{kurt}(s_m) h_{km}^* h_{lm} \end{bmatrix} \\ &= \begin{bmatrix} h_{11} & \dots & h_{1m} \\ \vdots & \ddots & \vdots \\ h_{m1} & \dots & h_{mm} \end{bmatrix}^H \\ &= \mathbf{H} \mathbf{\Lambda}_{Q^s} \mathbf{H}^H \end{aligned} \quad (5.42)$$

Im Fall von geweißten Signalen \mathbf{z} wird die Matrix \mathbf{H} durch die Matrix $\mathbf{T}^H = \mathbf{V}\mathbf{H}$ ersetzt. Dann gilt:

$$\hat{\mathbf{Q}}_{kl}^z = \mathbf{T}^H \mathbf{\Lambda}_{Q^z} \mathbf{T}$$

Umstellen nach $\mathbf{\Lambda}_{Q^z}$ wegen $\mathbf{T}^{-1} = \mathbf{T}^H$ ergibt

$$\mathbf{\Lambda}_{Q^z} = \mathbf{T} \hat{\mathbf{Q}}_{kl}^z \mathbf{T}^H. \quad (5.43)$$

Gleichung (5.43) zeigt, daß die gesuchte Transformationsmatrix \mathbf{T} die Matrix der Eigenvektoren der Kumulantenmatrizen der geweißten Signale ist, da die Multiplikation mit dieser Eigenmatrix zur Diagonalisierung von $\hat{\mathbf{Q}}_{kl}^z$ führt. Es wäre theoretisch möglich, aus den geweißten Signalen nur eine Kumulantenmatrix zu bestimmen, deren Eigenvektoren dann das gewünschte Ergebnis liefern. Das Problem dabei ist, daß

sie nur dann eindeutig sind, wenn alle Eigenwerte verschieden sind (Cardoso (1999)). Da jedoch nicht vorausgesagt werden kann, für welche Kumulantenmatrix dies der Fall ist, diagonalisiert man alle Kumulantenmatrizen gleichzeitig, was natürlich nur annähernd geschehen kann. Der daraus entstandene Algorithmus wird auch JADE (*Joint Diagonalisation of Eigenmatrices*) genannt. Dabei wird folgendes Kriterium optimiert:

$$\hat{\mathbf{V}} = \arg \min_{\mathbf{T}} \sum_{kl} \text{Off}(\mathbf{T} \hat{\mathbf{Q}}_{kl}^z \mathbf{T}^H) \quad (5.44)$$

mit $\text{Off}(\mathbf{B}) = \sum_{i \neq j} |b_{ij}|^2$. Diese Bedingung diagonalisiert die Kumulantenmatrizen. Zur gemeinsamen Diagonalisierung wird der von Cardoso und Souloumiac (1996) eingeführte Jacobi-Algorithmus genutzt. Alternativ lassen sich die Kreuzkumulanten zu einem Energiemaß zusammenfassen, das dann wie zuvor für die Negentropie oder die gemeinsame Information beschrieben minimiert wird (vgl. Mejuto *et al.* (2000) oder Dapena *et al.* (2001)).

5.3 Kontrastfunktionen auf Basis der Statistik zweiter Ordnung (SOBI)

So wie (5.3) die Grundlage für die Verfahren im letzten Abschnitt war, basiert dieses Kapitel auf den linearen Gleichung (5.6) und (5.5).

Erweitert man diese Gleichungen auf die Vektoren der breitbandigen rekonstruierten Signale, führt dies darauf, daß die bekannten Kovarianzmatrizen unter der Annahme mittelwertfreier Signale zu allen Zeitpunkten t und allen Verzögerungen τ Diagonalmatrizen sein sollen:

$$\mathbf{R}_{\mathbf{y}}^b(t, \tau) = E\{\mathbf{y}^b(t)\mathbf{y}^{bT}(t + \tau)\} = \mathbf{\Lambda}_{\mathbf{y}}^b(t, \tau).$$

Das setzt implizit voraus, daß die Elemente auf der Hauptachse nicht verschwinden, sonst erhält man Null-Matrizen. Die Hauptachsenelemente sind die Autokovarianzfunktionen der Signale. Damit sie für $\tau \neq 0$ nicht verschwinden, dürfen die Signale nicht (zeitlich) unkorreliert oder spektral weiß sein. Davon wird bei Motorgeräuschen und Sprachsignalen ausgegangen. Dafür kann die Forderung, daß die Signale nicht normalverteilt sein dürfen, fallengelassen werden.

Im skalaren Fall ist die Diagonalisierung mehrerer Kovarianzmatrizen $\mathbf{R}_{\mathbf{y}}^b(t, \tau)$ mit einer einzigen Entmischungsmatrix \mathbf{W} hinreichend für eine Identität zwischen den rekonstruierten und den Quellensignalen. Erstmals wurde ein Algorithmus für verschiedene Verzögerungen von Molgedy und Schuster (1994) vorgeschlagen, der im AMUSE-Algorithmus von Hyvärinen *et al.* (2001) mündete. Weitere bekannte Algorithmen sind SOBI (second order blind source separation) von Belouchrani *et al.* (1997) bzw. TDSEP (Temporal difference source separation) von Ziehe und Müller (1998). Sie bestehen aus zwei Stufen. Zunächst werden die Signale mit einer PCA (5.11) geweißt. Danach wird nur noch eine Orthogonalmatrix gesucht, die die Elemente aller Matrizen, die nicht auf der Hauptachse liegen (Nichtdiagonalelemente),

minimiert; d.h. es wird eine Verbunddiagonalisierung ausgeführt:

$$\mathbf{T} = \arg \min_{\mathbf{T}} \sum_{\tau} \sum_{i \neq j} \left| (\mathbf{T} \mathbf{R}_z(\tau) \mathbf{T}^H)_{ij} \right|^2. \quad (5.45)$$

Verfahren für instationäre Signale wurden von Matsouka *et al.* (1995) eingeführt. Das zu minimierende Fehlermaß ist sehr ähnlich:

$$\mathbf{T} = \arg \min_{\mathbf{T}} \sum_t \sum_{i \neq j} \left| (\mathbf{T} \mathbf{R}_z(t) \mathbf{T}^H)_{ij} \right|^2. \quad (5.46)$$

Zu dessen Minimierung können grundsätzlich die gleichen Algorithmen wie für die zeitverzögerte Dekorrelation verwendet werden. Speziell wurden u.a. ein LMS-Algorithmus von Parra *et al.* (1998), ein RMS-Ansatz von Jones (1999) oder ein Algorithmus auf Basis der Eigenwertzerlegung von Pham und Cardoso (2000) vorgeschlagen.

Im konvolutiven Fall ist diese Verbunddiagonalisierung nicht ohne weiteres möglich. Um das zu veranschaulichen, seien die Signale stationär und $\mathbf{R}_s^b(\tau)$, $\mathbf{R}_x^b(\tau)$ und $\mathbf{R}_y^b(\tau)$ die Kovarianzmatrizen der Quellen-, Observations- und rekonstruierten Signale. Damit folgt aus der Bedingung, daß die Quellensignale statistisch unabhängig sein sollen,

$$\mathbf{R}_s^b(\tau) = \mathbf{\Lambda}_s^b(\tau) = \text{diag}[\lambda_1, \dots, \lambda_M]. \quad (5.47)$$

Aus der Annahme für die akustische Überlagerung aus Kap. 4.1 ergibt sich

$$\begin{aligned} E\{\mathbf{x}^b(t) \mathbf{x}^{bT}(t + \tau)\} &= E\left\{ \sum_k \mathbf{h}_i^b(k) \mathbf{s}^b(t - k) \sum_k \mathbf{s}^{bT}(t - \tau - k) \mathbf{h}_j^{bT}(k) \right\} \\ &= E\left\{ \sum_k \sum_l \mathbf{h}_i^b(k) \mathbf{s}^b(t - k) \mathbf{s}^{bT}(t - \tau - l) \mathbf{h}_j^{bT}(l) \right\} \\ &= \sum_k \sum_l \mathbf{h}_i^b(k) \mathbf{R}_s^b(k - \tau - l) \mathbf{h}_j^{bT}(l) \end{aligned}$$

$$\begin{aligned} \Rightarrow \quad \mathbf{R}_x^b(\tau) &= \sum_k \sum_l \mathbf{H}^b(k) \mathbf{R}_s^b(k - \tau - l) \mathbf{H}^{bT}(l) \\ &= \mathbf{H}^b * \mathbf{\Lambda}_s^b(\tau) * \mathbf{H}^{bT}. \end{aligned} \quad (5.48)$$

Und mit analoger Betrachtung ergibt sich aus der Rekonstruktion mit FIR-Filtern nach Kap. 5.1.2 für die Entmischung

$$\mathbf{R}_y^b(\tau) = \mathbf{W}^b * \mathbf{R}_x^b(\tau) * \mathbf{W}^{bT}. \quad (5.49)$$

Aus (5.49) wird ersichtlich, daß man die Filterkoeffizienten nicht unabhängig voneinander bestimmen kann, weil sie auf mehrere Gleichungen einen unterschiedlichen Einfluß haben. Transformiert man (5.48) und (5.49) in den z -Bereich, ergibt sich mit dem Faltungstheorem für Vektoren im z -Bereich für die Mischung

$$\mathbf{R}_x(z) = \mathbf{H}(z) \mathbf{\Lambda}_s(z) \mathbf{H}^H(z)$$

und die Entmischung

$$\mathbf{R}_y(z) = \mathbf{W}(z)\mathbf{R}_x(z)\mathbf{W}^H(z) = \mathbf{\Lambda}_y(z).$$

Damit kann für jedes z die Übertragungsfunktion der Entmischungsfiler einzeln bestimmt werden. Allerdings steht dazu nur noch eine Gleichung zur Verfügung. Das heißt, die Diagonalisierung von $\mathbf{R}_y(z)$ entspricht einer PCA, d.h. einer klassischen Dekorrelation in jedem Frequenzbereich. Eine PCA allein ist jedoch nicht hinreichend für eine Trennung der Signale, weil sie, wie in Kap. 5.1.4 beschrieben, noch immer Mehrdeutigkeiten enthält¹⁸.

Auf dieses Problem wurde schon von Tong und Soon (1991) und Weinstein *et al.* (1993) hingewiesen¹⁹. Dennoch wurden einige Algorithmen auf dieser Basis²⁰ veröffentlicht. Die erfolgreiche Trennung durch diese Algorithmen führen die Autoren selbst darauf zurück, daß durch den Online-Charakter immer wieder neue Kovarianzmatrizen geschätzt werden, was bei instationären Prozessen zur gleichzeitigen Diagonalisierung von Kovarianzmatrizen unterschiedlicher Zeitpunkte führt. Damit sind diese Verfahren äquivalent zu den im folgenden eingeführten Verfahren von Parra und Spence (2000).

Das Problem läßt sich allgemein nur lösen, indem man zusätzliche Bedingungen aus der Instationarität der Signale generiert. Nimmt man instationäre Signale an²¹, unterscheiden sich die $\mathbf{R}_y^b(t, \tau)$ für verschiedene t . Durch eine gleichzeitige Diagonalisierung wird nun eine Entmischung erreicht. Dieses Verfahren wurde auf der theoretischen Grundlage von Yellin und Weinstein (1994) durch Parra und Spence (2000) popularisiert, die eine Diagonalisierung mit einem LMS-Verfahren vorschlugen. Zeitgleich oder vorher entwickelten auch Chan *et al.* (1996), Fabry *et al.* (1998), Servièrè (1998), Murata *et al.* (1998), Mejuto und Principe (1999), Wu und Principe (1999) Algorithmen auf dieser Basis, wobei die letztgenannten eine Verbunddiagonalisierung von Cardoso und Souloumiac (1996) verwendeten. Sie testeten ihre Algorithmen nur mit theoretischen Signalen, während Parra und Spence (2000) die Anwendbarkeit auch mit realen Signalen zeigten. Erweiterungen und Modifikationen

¹⁸Anschaulich kann man sich das klarmachen, indem man die Matrizen $\mathbf{R}_y(z)$ auf dem Einheitskreis betrachtet. Bei ihnen stehen auf der Hauptdiagonalen die Autoleistungsdichtespektren und die anderen Elemente sind die Kreuzleistungsdichtespektren. Damit werden bei der zeitverzögerten Dekorrelation lediglich die Kreuzleistungsdichten minimiert.

¹⁹Für den Spezialfall minimalphasiger Mischungen im 2×2 -Fall wurde durch Lindgren und Broman (1998) gezeigt, daß eine einfache Dekorrelation ausreichend sein kann. Dieser Spezialfall tritt bei realen akustischen Mischungen jedoch nur auf, wenn alle Echos schwächer als der Direktschall sind und sich jede Quelle in der Nähe „ihres Mikrofons“ befindet. Da dies allgemein nicht garantiert werden kann, werden die darauf basierenden Algorithmen nicht weiter diskutiert.

²⁰z.B. von Schobben und Sommen (1998) und Ehlers und Schuster (1997)

²¹Das erscheint zunächst schwierig, weil die Erwartungswerte bei Vorhandensein nur einer Realisierung des Zufallsprozesses traditionell aus dem unendlich ausgedehnten zeitlichen Mittel geschätzt werden. Das setzt eine Ergodizität und damit eine Stationarität des Prozesses voraus. Diese Schwierigkeit wird dadurch gelöst, daß die Signale als kurzzeitstationär angenommen werden, so daß die Erwartungswerte hinreichend genau geschätzt werden können. Langfristig müssen sich die statistischen Kennwerte jedoch ändern, was bei amplitudenmodulierten Signalen, wie Sprache oder Maschinengeräuschen, auch der Fall ist. Für eine ausführlichere Diskussion dieser Vorgehensweise, siehe Murata *et al.* (1998).

erfolgten seitdem von zahlreichen Autoren, z.B. wurde von Asano *et al.* (2000) eine Reduktion auf einen Unterraum („subspace“) eingeführt, die zu einer Rauschunterdrückung führt, wenn mehr Mikrofone als Quellen zur Verfügung stehen.

Damit ergibt sich nach der z -Transformation des Systems

$$\mathbf{R}_y(z, t) = \mathbf{W}(z)\mathbf{R}_x(z, t)\mathbf{W}^H(z) = \mathbf{\Lambda}_y(z, t).$$

Tastet man die z -Ebene äquidistant entlang des Einheitskreises ab, ist dies äquivalent mit der STFT und es ergibt sich, unter der Voraussetzung, daß die FFT-Länge größer als die Filterlänge ist (Verhinderung einer zyklischen Faltung, siehe Oppenheim und Schaffer (1995)),

$$\mathbf{R}_y(f, t) = \mathbf{W}(f)\mathbf{R}_x(f, t)\mathbf{W}^H(f) \Rightarrow \mathbf{R}_y(t) = \mathbf{W}\mathbf{R}_x(t)\mathbf{W}^H.$$

Dies entspricht der Darstellung im Frequenzbereich, die bereits für die Beamformer und die nichtlineare Dekorrelation angewendet wurde, weshalb auch hier der obere Index für die Frequenz weggelassen wird, siehe Kap. 4.1.

Für jeden Frequenzabschnitt hat man somit mehrere linear unabhängige Gleichungen, deren gleichzeitige Diagonalisierung eine hinreichende Bedingung für die Entmischung ist. Für die Methode der gleichzeitigen Diagonalisierung bieten sich alle Verfahren an, die schon für die skalaren Mischungen eingeführt wurden.

Im einfachsten Fall werden nur zwei Matrizen verwendet, für die von Molgedy und Schuster (1994) eine analytische Lösung angegeben wurde, indem sie folgendes Eigenwertproblem aufstellten:

$$\begin{aligned} \mathbf{R}_{t_1}^{-1} &= \mathbf{H}^{-H} \mathbf{\Lambda}_{t_1}^{-1} \mathbf{H}^{-1} \\ \mathbf{R}_{t_2} \cdot \mathbf{R}_{t_1}^{-1} &= \mathbf{H} \mathbf{\Lambda}_{t_2} \mathbf{H}^H \cdot \mathbf{H}^{-T} \mathbf{\Lambda}_{t_1}^{-1} \mathbf{H}^{-1} \\ &= \mathbf{H} \mathbf{\Lambda}_{t_2} \mathbf{\Lambda}_{t_1}^{-1} \mathbf{H}^{-1} \\ \mathbf{R}_{t_2} \cdot \mathbf{R}_{t_1}^{-1} \cdot \mathbf{H} &= \mathbf{H} \mathbf{\Lambda}_{t_2} \mathbf{\Lambda}_{t_1}^{-1}. \end{aligned} \quad (5.50)$$

Da das Produkt zweier diagonaler Matrizen ebenfalls diagonal ist, stellt (5.50) ein Eigenwertproblem dar. Durch die Eigenwertzerlegung von $\mathbf{R}_{t_2} \cdot \mathbf{R}_{t_1}^{-1}$ läßt sich \mathbf{H} bestimmen und das Trennungsproblem durch Invertierung von \mathbf{H} lösen. Das Problem dieser Methode liegt in der Wahl geeigneter Zeitpunkte t_1 und t_2 . Wenn sich die Kovarianzen zu diesen Zeitpunkten nicht hinreichend unterscheiden, degeneriert das Optimierungsproblem zu einer reinen PCA.

Sollen mehr als zwei Matrizen gleichzeitig diagonalisiert werden, bieten sich eine analytische Lösung auf Grundlage der Eigenwertzerlegung und eine iterative Lösung an (Gradientenabstiegsverfahren). Das Gradientenabstiegsverfahren, welches Parra und Spence (2000) benutzen, minimiert die Nichtdiagonalelemente der Matrizen. Dazu wird die Frobenius-Norm der Kovarianzmatrizen abzüglich der geschätzten Leistung der Quellensignale ($\mathbf{\Lambda}_y(t) = \text{diag}[\lambda_1(t), \dots, \lambda_N(t)]$) minimiert:

$$J = \sum_{t=1}^T \|\mathbf{W}\mathbf{R}_x(t)\mathbf{W}^H - \mathbf{\Lambda}_y(t)\|^2 = \sum_{t=1}^T \|\mathbf{J}_t\|_F^2. \quad (5.51)$$

Die zu variierenden Größen sind r_{nm} und $\lambda_n(t)$. Zur Optimierung werden wieder die komplexen Gradienten nach diesen Größen (siehe Kap. 4.3.1) verwendet. Zunächst wird die Leistung mit der leicht zu zeigenden Eigenschaft der Frobenius-Norm $\|\mathbf{M}\|_F^2 = \text{tr}(\mathbf{M}\mathbf{M}^H)$ und der konjugiert-komplexen Symmetrie vereinfacht.

$$\begin{aligned} J &= \sum_{t=1}^T \text{tr} \left((\mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H - \Lambda_{yt})(\mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H - \Lambda_{yt})^H \right) \\ &= \sum_{t=1}^T \left(\text{tr}(\mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H \cdot \mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H) - \text{tr}(\Lambda_{yt} \cdot \mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H) \right. \\ &\quad \left. - \text{tr}(\mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H \Lambda_{yt}^H) + \text{tr}(\Lambda_{yt} \cdot \Lambda_{yt}^H) \right) \end{aligned}$$

Für den komplexen Gradienten nach $\Lambda_{yt}(t)^*$ dürfen nur die Elemente auf der Hauptdiagonalen verwendet werden, weil nur diese durch die Least-squares-Methode manipuliert werden sollen. Damit ergibt sich mit $\nabla_{\mathbf{B}^*} \text{tr}(\mathbf{A}\mathbf{B}^H) = \mathbf{A}$ und $\nabla_{\mathbf{A}^*} \text{tr}(\mathbf{A}\mathbf{A}^H) = \mathbf{A}$

$$\begin{aligned} \nabla_{\text{diag}(\Lambda_{nn,t}^*)} J &= \text{diag} \left[\nabla_{\Lambda_t^*} \left(-\text{tr}(\mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H \Lambda_{yt}^H) + \text{tr}(\Lambda_{yt} \Lambda_{yt}^H) + \text{const} \right) \right] \\ &= -\text{diag} \left[\mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H - \Lambda_{yt} \right] = \text{diag}[\mathbf{J}_t]. \end{aligned}$$

Für die Entmischungsmatrix ergibt sich

$$\begin{aligned} \nabla_{\mathbf{W}^*} J &= \sum \nabla_{\mathbf{W}^*} \text{tr}(\mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H \cdot \mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H) - \nabla_{\mathbf{W}^*} \text{tr}(\Lambda_{yt} \cdot \mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H) \\ &\quad - \nabla_{\mathbf{W}^*} \text{tr}(\mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H \Lambda_{yt}^H) \\ &= \sum 2 \cdot \mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H \mathbf{W}\mathbf{R}_{xt} - \Lambda_{yt} \mathbf{W}\mathbf{R}_{xt} - \Lambda_{yt} \mathbf{W}\mathbf{R}_{xt} \\ &= \sum 2 \cdot (\mathbf{W}\mathbf{R}_{xt}\mathbf{W}^H - \Lambda_{yt}) \mathbf{W}\mathbf{R}_{xt} \\ &= \sum 2\mathbf{J}_t \mathbf{W}\mathbf{R}_{xt} \end{aligned}$$

mit der Eigenschaft eines Matrizenprodukts, daß die Spur bei Veränderung der Reihenfolge unverändert bleibt. Alternativ kann man diese Minimierung über einen Algorithmus erreichen, bei dem durch wiederholtes Anwenden von Jacobi-Drehungen mehrere Matrizen durch einen orthogonalen Operator diagonalisiert werden. Der Algorithmus gleicht dem im letzten Kapitel beschriebenen JADE-Algorithmus, bei dem ebenfalls mehrere Matrizen gleichzeitig diagonalisiert werden mußten, und wurde von Cardoso und Souloumiac (1996) ausführlich beschrieben.

Die Schätzung der Kovarianzmatrizen der schmalbandigen Signale kann nach der Methode von Welch (vgl. Kammeyer und Kroschel (1998)) leicht über eine Mittelung der Kurzzeitspektren erfolgen (5.52). Dabei sind die Parameter T und N von besonderer Bedeutung, weil einerseits garantiert sein muß, daß es sich um eine konsistente Schätzung handelt (großes N für eine lange Mittelung), und andererseits N klein sein muß, weil es sich um instationäre Signale handelt. Die Parameter müssen also in Abhängigkeit von der Stationaritätslänge der Signale gewählt werden: Bei Sprache etwa 100 ms und bei Maschinensignalen etwa die Zeit einer Umdrehungszeit.

$$\hat{\mathbf{R}}_x(f, t) = \frac{1}{N} \sum_{n=0}^N \mathbf{x}(f, t + nT) \mathbf{x}^H(f, t + nT) \quad (5.52)$$

5.4 Behandlung des Permutations- und Skalierungsproblems

Wie in Kap. 5.1.3 angegeben, stellen die Permutations- und Skalierungsmehrdeutigkeiten der ICA ein Problem dar. Dies wurde von Smaragdis (1998) zunächst als Bagatelle angesehen, stellt aber heute neben den langen Impulsantworten realistischer Räume das Haupthindernis für die praktische Anwendbarkeit der konvolutiven ICA dar.

5.4.1 Allgemeine Methoden zur Behandlung des Permutationsproblems

Die bisher vorgeschlagenen Ansätze zur Lösung des Permutationsproblems lassen sich in drei Gruppen unterteilen: Maximierung der Ähnlichkeit zwischen den Frequenzbändern, Begrenzung der Länge der Entmischungfilter und geometrische Verfahren.

Die Verfahren, die die Ähnlichkeit zwischen den schmalbandigen Signalen unterschiedlicher Frequenzbänder maximieren, basieren auf der Annahme, daß die Amplitudenmodulation aller schmalbandigen Teilsignale ähnlich ist. Motiviert wird diese Annahme dadurch, daß beispielsweise eine Lautstärk modulation auf alle Frequenzbänder wirkt. Eingeführt bzw. verwendet wird dieses Verfahren von Murata *et al.* (1998), Anemüller und Kollmeier (2000), Davies (2001) u.a., wobei die Darstellung der Amplitudenmodulation sehr unterschiedlich ist.

Berechnet man die Amplitudenmodulation der schmalbandigen, rekonstruierten Teilsignale beispielsweise mit der Einhüllenden nach Murata *et al.* (1998)

$$\mathcal{E}\{y_i^f(t_s)\} = \frac{1}{2M} \sum_{t'_s=t_s-M}^{t_s+M} |y_i^f(t'_s)|,$$

kann man die Ähnlichkeit durch

$$\text{sim}(y_i^{f1}, y_j^{f2}) = \frac{\mathcal{E}\{y_i^{f1}\} \cdot \mathcal{E}\{y_j^{f2}\}}{\|y_i^{f1}\| \cdot \|y_j^{f2}\|} \quad (5.53)$$

mit dem inneren Produkt der Einhüllenden

$$\mathcal{E}\{y_i^{f1}\} \cdot \mathcal{E}\{y_j^{f2}\} = \sum_{\forall t_s} \mathcal{E}\{y_i^{f1}(t_s)\} \cdot \mathcal{E}\{y_j^{f2}(t_s)\} \quad (5.54)$$

und

$$\|\mathcal{E}\{y_i^{f1}\}\| = \sqrt{\mathcal{E}\{y_i^{f1}\} \cdot \mathcal{E}\{y_i^{f1}\}}$$

ausdrücken. Da hier schmalbandige Signale unterschiedlicher Frequenz behandelt werden, wird der obere Index für die Frequenz wieder benutzt.

Murata *et al.* (1998) sortieren nun die Frequenzbänder nach der Ähnlichkeit der geschätzten schmalbandigen Signale gemäß (5.53). Beginnend mit dem Frequenzband f_1 , in dem sich die Signale am meisten unterscheiden, wird nun in jedem weiteren Frequenzband f_{k+1} jedes neue Signal $y_i^{f_{k+1}}$ mit der Einhüllenden der Summe der bereits zugeordneten Kanäle verglichen und entsprechend zugeordnet.

$$\text{sim} \left(y_i^{f_{k+1}}, \sum_{n=0}^k y_j^{f_n} \right) \rightarrow \max$$

Die Anzahl der Permutationsmöglichkeiten liegt bei $N!$ pro Frequenzband für den $N \times N$ -Fall. Damit erklärt sich auch, warum dieses Verfahren auf den 2×2 -Fall begrenzt ist.

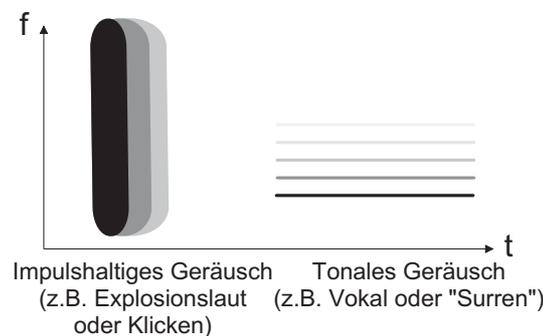


Abb. 5.4: Schematisches Spektrogramm eines Störsignals, bei dem nicht alle Frequenzbänder miteinander korreliert sind.

Dieses Verfahren hat den Nachteil, daß die Einhüllenden in weit entfernten Frequenzbändern verglichen werden, deren Amplitudenmodulation nur schwach oder gar nicht gekoppelt sein können, wie Abb. 5.4 illustriert. Sie stellt das schematische Spektrogramm eines Sprachsignals dar. Bei stimmlosen Explosivlauten (z.B. „p“) zum Zeitpunkt t_1 sind die Einhüllende korreliert, zum Zeitpunkt t_2 hingegen ist nur der untere Frequenzbereich angeregt, so daß die Amplitudenmodulation nicht einheitlich ist (vgl. Knorr (2002)). Ähnliches gilt für Störgeräusche, wenn sie eher ein tonales Geräusch abgeben (z.B. „Surren“).

Eine Alternative ist es, bei der kleinsten Frequenz anzufangen und mit (5.53) nur die benachbarten Bänder zu vergleichen. Hier tritt jedoch das Problem auf, daß, wenn man sich einmal „geirrt“ hat, die schmalbandigen Signale aller weiteren Frequenzbänder nicht richtig zugeordnet werden. Abb. 5.5²² stellt die schlechte Lösung der Permutationsprobleme durch die Korrelationsmethode dar: Während in Abb. 5.5a in allen Frequenzabschnitten die gleiche Permutation gewählt wurde, tritt bei der Korrelationsmethode in Abb. 5.5b ein Wechsel der Permutation auf. Die Konsequenz ist, daß die Signale weiter gemischt erscheinen, weil die höheren Frequenzen von dem einen, die tiefen jedoch von dem anderen Quellensignal kommen.

²²Die dargestellte Größe „frequenz aufgelöster SIR“ wird in Kap. 8.2 eingeführt. Sie gibt den Anteil der Störung, in diesem Fall des jeweils anderen Quellensignals, am rekonstruierten Signal an.

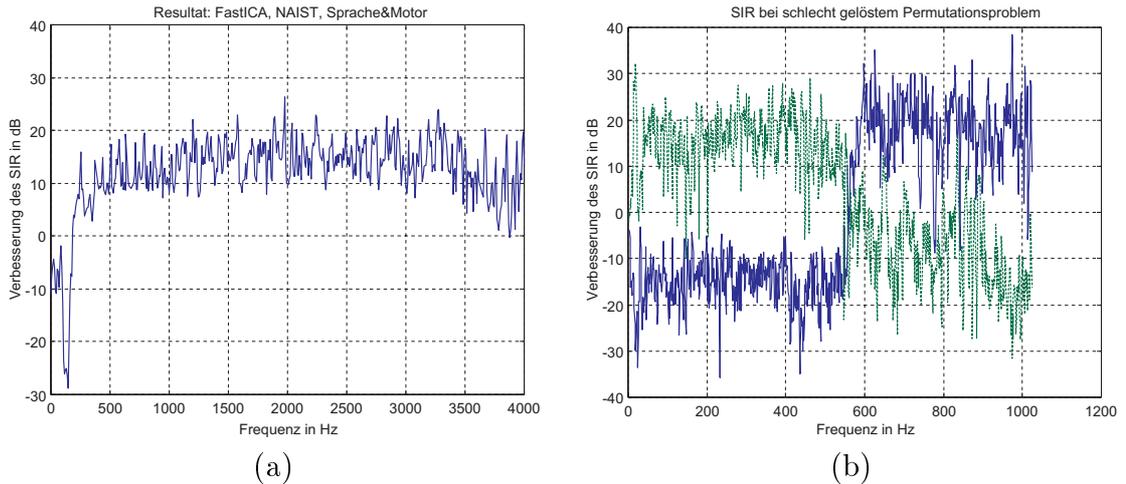


Abb. 5.5: Frequenzaufgelöster SIR eines FastICA-Algorithmus. Das Permutationsproblem ist in (a) mit dem geometrischen Verfahren gut und mit der Korrelationsmethode (b) schlecht gelöst. Der Gesamt-SIR ist in (b) sehr gering, obwohl in jedem Frequenzbereich eine gute Entmischung erzielt wurde.

Der Ansatz von Parra und Spence (2000) geht von einer anderen Annahme aus. Die Länge des Entmischungsfilters w_{ij}^b ist begrenzt, so daß seine Übertragungsfunktion „glatt“ sein muß. Denn ein Wechsel der Permutation führt zu einer zusätzlichen Sprungstelle und damit zu einer längeren Impulsantwort.

Projiziert man darum die gefundenen Entmischungsfiler in den Raum „kurzer Impulsantworten“ wird das Permutationsproblem gelöst. Praktisch geschieht das, indem man unter der Voraussetzung, daß die maximale Entmischungsfilerlänge viel kleiner als die Länge der STFT ist, durch die Projektion

$$\mathbf{P} = \mathbf{F} \cdot \mathbf{Z} \cdot \mathbf{F}^{-1}$$

mit der DFT-Matrix \mathbf{F} und einem Operator \mathbf{Z} , der die letzten $T - K$ Werte von w_{ij}^b zu Null setzt. T ist die Länge der DFT.

Der Vorteil dieser Methode liegt in ihrer Einfachheit. Sie setzt allerdings voraus, daß bereits vor ihrer Anwendung die Permutationen in den meisten Frequenzbändern gleich sind. Denn Permutationen werden nicht gewechselt, sondern „weg geglättet“. Ein größerer Nachteil liegt in der Begrenzung der Filterlänge. Florenco *et al.* (2001) haben darum einen mehrstufigen Algorithmus mit wachsender Filterlänge vorgeschlagen. Aber auch dies ändert das Grundproblem nicht entscheidend: Ein realistischer Raum hat Impulsantworten mit mehreren tausend Koeffizienten, so daß eine entsprechend langer DFT benötigt wird. Sie erfordert nicht nur viel Rechenzeit, sondern läßt bei einer begrenzten Datenmenge auch die Schätzung der Kovarianzmatrizen unzuverlässig werden, weil in jedem Frequenzabschnitt immer weniger Daten zur Verfügung stehen (Araki *et al.* (2001)).

Die geometrischen Verfahren wurden von Sawada *et al.* (2002b), Saruwatari *et al.* (2001) und Kurita *et al.* (2000) eingeführt und benutzen die später in (Kap. 6.1) noch genauer besprochene Identität zwischen adaptiven Beamformern und der konvolutiven ICA. Jeder Kanal stellt einen schmalbandigen, adaptiven Beamformer dar, der

sich auf eine Quelle adaptiert hat. Sein Beampattern hat also ein Maximum in Richtung dieser Quelle und Nullstellen in den Richtungen der anderen Quellen. Durch die Orthogonalisierung wird ausgeschlossen, daß sich alle Kanäle auf die gleiche Quelle adaptieren. Abb. 5.5b zeigt, daß das geometrische Verfahren zu einer konsistenten Zuordnung der rekonstruierten Signale in den Teilbändern gelangt. Lediglich für extrem kleine Frequenzen kann die Phase nicht genau ermittelt werden, weil sie sehr klein ist und sich Rundungsfehler etc. bemerkbar machen.

Das Beampattern jedes Kanals kann man aus den Erkenntnissen der vorangegangenen Kapitel berechnen. Nach (3.20) gilt, wenn sich die Mikrofone in einem Linienarray (Abb. 4.1) an den Positionen \vec{x}_n befinden und bei \vec{x}_0 der Arraymittelpunkt liegt,

$$W_n(f, \vec{\xi}) = \sum_k w_{nk}^f \cdot e^{j2\pi f(x_n - x_0)\vec{\xi}/c}. \quad (5.55)$$

Da mit (5.55) die Wirkung von \mathbf{w}_n^f auf schmalbandige Signale aus Richtung $\vec{\xi}$ beschrieben wird, kann man die Nullrichtungen bestimmen und die Kanäle zusammenfassen, die in der jeweiligen Frequenz auf die gleiche Richtung eine Null gelegt haben. Bei einem zweikanaligen System ergibt sich

$$\begin{aligned} W_{neu_1}^f &= \min[\arg \min_{\vec{\xi}} W_1 \quad \arg \min_{\vec{\xi}} W_2] \\ W_{neu_2}^f &= \max[\arg \min_{\vec{\xi}} W_1 \quad \arg \min_{\vec{\xi}} W_2]. \end{aligned}$$

Für Mischungen mit mehr als zwei Kanälen ist auch dieses Verfahren nicht einsetzbar. Weiterhin ist es sehr aufwendig, weil die W_n für alle $\vec{\xi}$ berechnet werden müssen. Von Sawada *et al.* (2002b) wurde darum eine schnellere Berechnung auf der gleichen Grundlage eingeführt.

Durch die Beziehung $e^{j\omega\tau}$ in (5.55) wird die Phasendifferenz im unteren Frequenzbereich sehr klein, so daß diese Verfahren für niedrige Frequenzen nicht anwendbar sind. Hier muß eine andere Methode, z.B. die Korrelation zwischen den Frequenzbändern verwendet werden.

Mit dieser Einschränkung sind die geometrischen Methoden die derzeit besten. Sie erreichen im zweikanaligen Fall auch bei realistischen Raumimpulsantworten eine optimale Unterdrückung des Permutationsproblems in einer akzeptablen Zeit.

5.4.2 Ausnutzung der Eigenschaften der Motorsignale

Zusätzlich zu den bisher beschriebenen Verfahren zur Lösung des Permutationsproblems haben Knaak und Filbert (2002) vorgeschlagen, die Eigenschaften des Motorsignals auszunutzen, um die Komponenten der Vektoren \mathbf{y} zu finden, die das Nutzsignal bilden. Zunächst werden die Eigenschaften, die dazu genutzt werden, genauer dargestellt.

Nutzbare Eigenschaften des Motorsignals

Das Motorsignal s_1^b ist als Signal einer rotierenden Maschine annähernd T_0 -periodisch. Es hat damit ein Linienspektrum, d.h. die Leistung ist auf die Drehzahlharmonischen konzentriert, und es läßt sich durch eine Fourierreihe (5.56) beschreiben. c_k gibt darin die Stärke der k -ten Harmonischen an. Das Leistungsdichtespektrum eines Fensterhebermotors mit der Drehfrequenz 80 Hz ist in Bild 5.6a dargestellt.

$$s_1^b(t) = \sum_{-\infty}^{+\infty} c_k \cdot e^{j2\pi k f_0 t} \quad \text{mit} \quad f_0 = \frac{1}{T_0} \quad (5.56)$$

Durch die Verarbeitung im Frequenzbereich werden nur schmalbandige Signalanteile

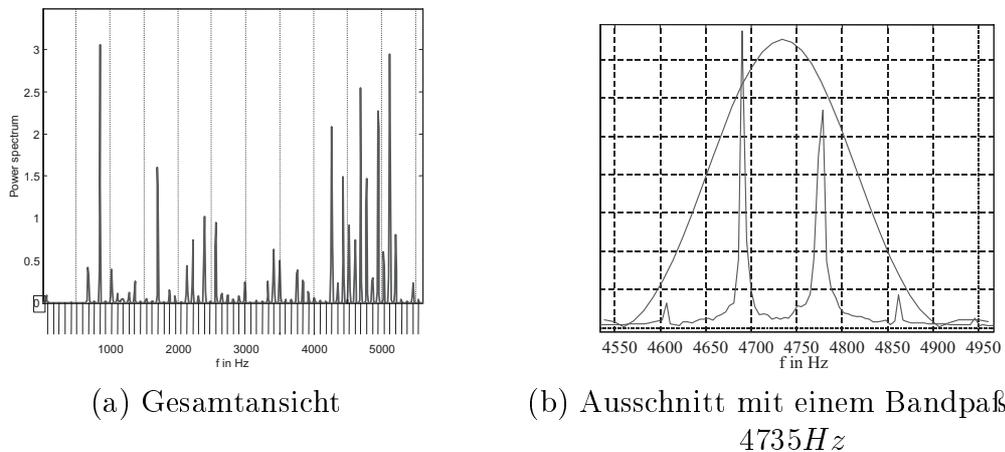


Abb. 5.6: Spektrum eines Fensterhebermotors mit den Drehzahlharmonischen (a - Gesamtansicht, b - Ausschnitt)

verwendet, die durch eine STFT entstanden sind. Sie wird hier als Filterbank mit analytischen Bandpässen angesehen. Unter Verwendung des Fensters $h(nT)$ lautet die bekannte Gleichung mit $f_x = \frac{m}{NT}$ und $t = nT$

$$s_1(t) = \sum_{n=0}^{N-1} s_1^b(nT + k\Delta T) \cdot h(nT) \cdot e^{-j2\pi \frac{nm}{N}}$$

In der Filterbank-Interpretation ist $h(nT) \cdot e^{-j2\pi \frac{nm}{N}}$ die Impulsantwort des asymmetrischen Bandpasses. Sei $H(f)$ die Übertragungsfunktion des durch das Fenster entstandenen Tiefpasses, so gilt

$$\tilde{H}(f) = H(f - f_x) = 0 \quad \forall f \leq 0.$$

und \tilde{H} ist ein analytischer Bandpaß mit der Mittenfrequenz f_x und der normierten Bandbreite Ω . Ein beispielhafter Bandpaß ist in Bild 5.6b angegeben.

Das gefilterte Signal $s(t)$ kann unter Annahme idealer Sperreigenschaften des

Bandpasses wie in (5.57) beschrieben werden.

$$\begin{aligned}
 s_1(t) &= \sum_{k \in K} \hat{c}_k \cdot e^{j2\pi k f_0 t} \\
 \text{mit } K &= \{\mathbf{N} \cap k f_0 [\in f_x(1 - \Omega), f_x(1 + \Omega)]\} \\
 \hat{c}_k &= \tilde{H}(k f_0) \cdot c_k
 \end{aligned} \tag{5.57}$$

Die Harmonischen im Sperrbereich fallen weg, im Durchlaßbereich werden sie modifiziert.

Man kann leicht zeigen, daß die Modulationsfunktion des so entstandenen Signals periodisch ist und die Periode der Drehfrequenz f_0 hat. Für ein analytisches Signal beschreibt die Einhüllende $A(t)$ die Amplituden- und $\Phi(t)$ die Frequenzmodulation, vgl. Oppenheim und Schaffer (1995). Damit gilt

$$s_1(t) = A(t) \cdot e^{j(2\pi f_x t + \Phi(t))}.$$

Weiterhin gilt

$$\begin{aligned}
 A^2(t) &= s_1 \cdot s_1^* = \sum_{k_1 \in K} \hat{c}_{k_1} \cdot e^{j2\pi k_1 f_0 t} \\
 &\quad \cdot \sum_{k_2 \in K} \hat{c}_{k_2}^* \cdot e^{-j2\pi k_2 f_0 t} \\
 &= \sum_{k_1} \sum_{k_2} \hat{c}_{k_1} \hat{c}_{k_2}^* e^{j2\pi f_0 t (k_1 - k_2)} \\
 &= \mathbf{1}^T \begin{bmatrix} |\hat{c}_1|^2 & \dots & \hat{c}_1 \hat{c}_K^* e^{j2\pi K f_0 t} \\ \vdots & \ddots & \vdots \\ \hat{c}_1^* \hat{c}_K e^{-j2\pi K f_0 t} & \dots & |\hat{c}_K|^2 \end{bmatrix} \cdot \mathbf{1}.
 \end{aligned}$$

Da es sich um eine hermitesche Matrix handelt, kann die $A^2(t)$ in einen Gleichanteil und Cosinus-förmige Anteile zerlegt werden.

$$\begin{aligned}
 A^2(t) &= \sum_{k=1}^K |\hat{c}_k|^2 + \\
 &\quad \sum_{l=0}^{K-1} \sum_{k=1+l}^{K-1-l} (\hat{c}_k^* \hat{c}_{k+l} e^{j2\pi l f_0 t} + \hat{c}_k \hat{c}_{k+1}^* e^{-j2\pi l f_0 t}) \\
 &= a_0^2 + \sum_l a_l \cdot \cos(2\pi l f_0 t + \phi_l)
 \end{aligned} \tag{5.58}$$

Der Gleichanteil ist gleich der Spur der Matrix und somit die quadratische Summe der im Signal noch enthaltenen Drehzahlharmonischen. Diese Äquivalenz ist nach dem Parsevalschen Theorem auch zu erwarten. Die Koeffizienten der Cosinusanteile ergeben sich aus den Summen der anderen Diagonalwerte der Matrix. In Abhängigkeit der Lage der Trägerfrequenz zu den Drehzahlharmonischen sind die Modulationen unterschiedlich stark ausgeprägt, jedoch immer vorhanden (vgl. Knaak und Filbert (2001b)), da nicht alle a_l null werden.

Bild 5.7b zeigt das Motorsignal, das mit dem in Bild 5.6b angegebenen Bandpaß gefiltert wurde. Obwohl durch die Verwendung eines asymmetrischen Bandpasses ein analytisches (und damit komplexes) Signal entsteht, wurde aus Gründen einer einfachen Visualisierung nur der Realteil gezeigt. Die periodische Modulation ist deutlich erkennbar. Nicht periodische Störsignale besitzen keine periodische Modulation ($k_2 = 0$), wie Bild 5.7a zeigt.

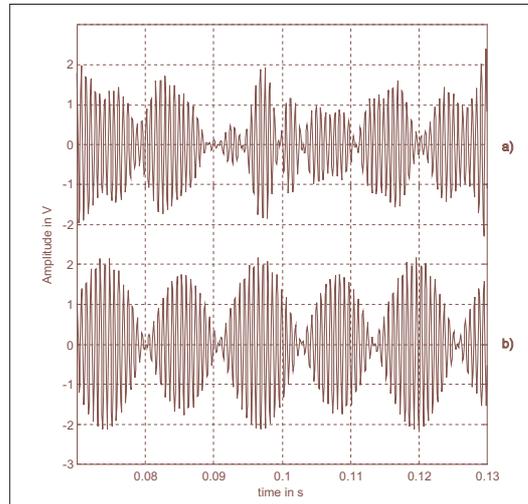


Abb. 5.7: Bandpaßgefilterte Quellensignale: (a) Störsignal,(b) Motorsignal

Algorithmus

Zunächst muß die Drehfrequenz f_0 aus den gemischten Daten ermittelt werden. Zu deren Bestimmung ist das Cepstrum (Definition siehe Oppenheim und Schafer (1995)) auch unter sehr gestörten Bedingungen zuverlässig, wie von Knaak *et al.* (2001) gezeigt wurde.

Darin wurden die Drehfrequenzen von Kleinmotoren²³ aus dem Schallsignal unter realen (gestörten) Bedingungen geschätzt. Dazu wurden das Schall- und das Stromsignal der Motoren, die in sechs Fehlerklassen eingeteilt waren, in einer Produktionshalle aufgenommen. Die Testbank (Abb. 5.8) war vom Fabrikbetrieb nur durch eine Plexiglasplatte getrennt, so daß der Störschallpegel in der Testbank bei 55-60 dB lag. Der Schallpegel der Motoren hatte einen Pegel von nur 45 dB.

Anhand der Stärke der dominierenden „Queffrency“ kann man die Zuverlässigkeit der Schätzungen bestimmen und ggf. unzuverlässige verwerfen. Abb. 5.9a zeigt, daß die Schätzzraten über 90% liegen und damit eine Schätzung fast immer auch unter schlechten Bedingungen möglich ist. In diesen Fällen wurde die geschätzte Drehzahl mit der aus dem Stromsignal bestimmten verglichen. Abb. 5.9b stellt die sich daraus ergebende Standardabweichung dar. Sie liegt unter 1%, so daß auch die Genauigkeit hinreichend ist.

²³Als Versuchsobjekt werden die Fensterhebermotoren genutzt, die in Kap. 7.1 genauer beschrieben sind.

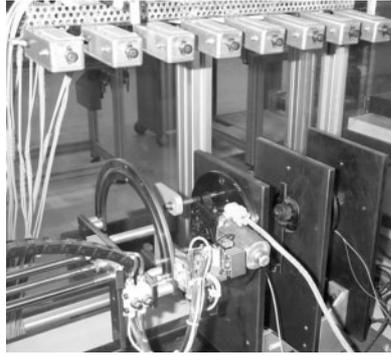


Abb. 5.8: Untersuchung von Elektromotoren in einer Produktionshalle

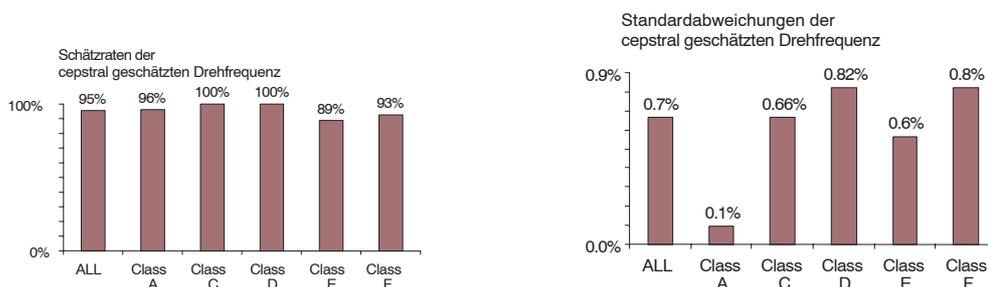


Abb. 5.9: Schätzzraten (a) und -genauigkeiten (b) der Drehzahl von gestörten Motorsignalen

Zur Bestimmung der Komponenten, die zum Motorsignal gehören, werden aus den Signalkomponenten die Einhüllenden $A(t)$ berechnet. Weisen sie einen spektralen Anteil bei f_0 auf, gehören sie zum Motorsignal. Für die Schätzung wird ein Bandpaß mit der Mittenfrequenz f_0 auf die Einhüllende angewendet. Übersteigt die Leistung dieses Signals einen Schwellwert, wird die Komponente als Teil des d.u.t.-Signals betrachtet und verwendet:

$$\frac{\|A_{y_k} * h_{f_0}\|_2}{\|A_{y_k}\|_2} > k \quad \Rightarrow \quad y_k = y_{opt}. \quad (5.59)$$

5.4.3 Skalierungsproblem

Die konvolutive ICA ist gegenüber einer Faltung invariant (vgl.Kap. 5.1.3). Damit ist es ohne weiteres nicht möglich, die durch den Raum entstandene Verzerrung rückgängig zu machen.

Die bisher vorgestellten Algorithmen benutzen eine PCA in jedem Frequenzbereich, so daß die Signale eine Einheitsvarianz erhalten. Das heißt für die breitbandigen Signale, daß sie ein Leistungsdichtespektrum identisch eins bekommen und zeitlich dekorreliert werden. Darum kann man entweder die PCA entfallen lassen oder eine Nachbearbeitung verwenden. Ein Wegfall der PCA ist bei vielen Algorithmen, wie der FastICA, gar nicht möglich, Bei anderen Algorithmen ist er nicht

sinnvoll, weil die PCA die Suche der Entmischungsvektoren auf den Einheitskreis beschränkt bleibt und somit die Konvergenz beschleunigt.

Darum wird eine Nachbearbeitung angewendet, die den Ausgangssignalen die Leistungsdichte der Observationssignale zuordnet. Das bedeutet den Verzicht auf die Entfaltung des Motorsignals, man rekonstruiert die Signale am Ort des Mikrofonarrays. Dies ist unproblematisch, weil die Verzerrung durch den Raum bei Sprache nicht besonders stark hörbar ist. In der Maschinendiagnose ist die Verzerrung dann nicht erheblich, wenn der Klassifikator ebenfalls mit Signalen, die an den Orten der Mikrofone aufgenommen wurden, trainiert wird.

Algorithmisch wird die Zuordnung in dieser Arbeit über den folgenden Gedankenansatz gelöst: Um aus den rekonstruierten Signalen wieder die Observationen zu erhalten, muß man letztere mit der Inversen der Mischungsmatrix multiplizieren:

$$\mathbf{x} = \mathbf{W}^+ \mathbf{y}.$$

Für jeden Kanal gilt, wenn \tilde{w}_{kn} ein Element der Inversen bezeichnet:

$$x_k = \tilde{\mathbf{w}}_k^H \mathbf{y} = \sum \tilde{w}_{kn} y_n.$$

Die Leistung von x_k entspricht der Varianz und die Kovarianzmatrix von \mathbf{y} ist die PCA eine Einheitsmatrix, so daß gilt:

$$\begin{aligned} \mathbb{E}\{|x_k|^2\} &= \mathbb{E}\{\tilde{\mathbf{w}}_k^H \mathbf{y} \mathbf{y}^H \tilde{\mathbf{w}}_k\} = \tilde{\mathbf{w}}_k^H \mathbb{E}\{\mathbf{y} \mathbf{y}^H\} \tilde{\mathbf{w}}_k \\ &= \tilde{\mathbf{w}}_k^H \mathbf{R}_y \tilde{\mathbf{w}}_k = \tilde{\mathbf{w}}_k^H \tilde{\mathbf{w}}_k = \sum |\tilde{w}_{kn}|^2. \end{aligned} \quad (5.60)$$

Wenn man diese Koeffizienten nur auf y_k anwendet,

$$\check{y}_k = \frac{1}{N} \sum |\tilde{w}_{kn}| \cdot y_k = \overline{|\tilde{w}_{kn}|} y_k,$$

dann ist die Leistung von \check{y}_k ungefähr gleich der halben Leistung des zugehörigen Observationssignals. Damit kann man die endgültige Entmischungsmatrix

$$\mathbf{W}_n = \text{diag}[\overline{|\tilde{w}_{1n}|}, \dots, \overline{|\tilde{w}_{Nn}|}] \cdot \mathbf{W} \quad (5.61)$$

angeben.

Kapitel 6

Vergleich adaptive Beamformer und blinde Quellentrennung

Zwischen der konvolutiven BSS und den adaptiven Beamformern besteht eine Identität, wie von Knaak und Filbert (2001a) und Araki *et al.* (2001) unabhängig voneinander nachgewiesen wurde. Die BSS besteht, wenn N Signale geschätzt werden, „nur“ aus N adaptiven Beamformern, deren maximale Durchlaßrichtung in der Richtung jeweils eines Signals liegt und die Nullen auf die anderen Signale gelegt werden. Die Verzögerungsvektoren werden implizit mitgeschätzt und müssen damit *nicht* vorgegeben werden.

Dies ist vom theoretischen Standpunkt bemerkenswert, weil die blinde Quellentrennung, aus der reinen Statistik bzw. der Theorie der neuronalen Informationsverarbeitung stammend, keine Aussagen über die Geometrie, Einfallrichtungen o.ä. macht oder solche Informationen nicht miteinbezogen werden können. Weiterhin erklärt die obige Identität die Wirkungsweise der BSS aus Sicht der mehrkanaligen Signalschätzung. Damit können die gut erforschten Gesetzmäßigkeiten der mehrkanaligen Signalverarbeitung genutzt werden, um z.B. die obere Grenzen der Störunterdrückung anzugeben etc.

In der praktischen Anwendung ist das wichtig, weil die Bedingungen, unter denen Mikrofon-Arrays optimal funktionieren, auch für die konvolute BSS gelten müssen. Die Arraygröße und der Mikrofonabstand müssen unter Berücksichtigung der gewünschten Ortsauflösung und des räumlichen Aliasing in Abhängigkeit vom untersuchten Frequenzbereich gewählt werden. Die Fragestellungen wurden in den ersten Arbeiten zur konvolutiven blinden Quellentrennung noch vernachlässigt.

Schließlich kann man auf dieser Identität auch robuste Algorithmen aufbauen, die die Vorteile beider Methoden nutzen. Solche Algorithmen wurden von Parra und Alvino (2001) und Knaak und Filbert (2001a) vorgeschlagen. Sie machten jedoch noch keine Aussagen über die Robustheit und den Anwendungsbereich, was erst von Knaak *et al.* (2003) untersucht wurde. Andere verbesserte Algorithmen, die auf der Identität aufbauen, wurden bereits in Kap. 5.4.3 vorgestellt. Sie berechnen aus den Entmischungsvektoren der BSS die Einfallrichtungen und lösen das Permutationsproblem, indem sie die Komponenten nach der Richtung sortieren. Derzeit sind sie

noch auf den Fall zweier Quellensignale beschränkt.

Zunächst wird die Identität formal gezeigt. Danach folgen die Ableitung des neuen Algorithmus und eine analytische Untersuchung seines Konvergenzgebietes.

6.1 Äquivalenz zwischen adaptiven Beamformern und konvolutiver BSS

Die Äquivalenz zwischen den beiden Konzepten wird zunächst für zwei Quellensignale und zwei Observationen gezeigt. Die Erweiterung auf N Kanäle ist dann trivial.

Lemma 5. *Die zweikanalige konvolute blinde Quellentrennung mit der Entmischungsmatrix \mathbf{W} besteht aus zwei adaptiven Beamformern mit den Steeringvektoren \mathbf{w}_{MV_1} und \mathbf{w}_{MV_2} , die auf jeweils eines der beiden Signale fokussiert sind, wobei exakte Verzögerungsvektoren angenommen werden.*

Beweis. Um den Beweis zu führen, muß gezeigt werden, daß die Steeringvektoren der beiden Beamformer den Vektoren der Entmischungsmatrix entsprechen.

Für den Steeringvektor des ersten Beamformers gilt, wie bereits hergeleitet nach (4.34) und (4.35)

$$\begin{aligned} p_1 &= \mathbf{w}_{MV_1}^H \hat{\mathbf{h}}_1 = 1 \\ p_2 &= \mathbf{w}_{MV_1}^H \hat{\mathbf{h}}_2 = 0. \end{aligned}$$

Fügt man den zweiten Beamformer, der auf die Quelle s_2 gerichtet ist, hinzu, gilt für diesen das gleiche und es ergibt sich:

$$[\mathbf{w}_{MV_1}^H \ \mathbf{w}_{MV_2}^H] \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Bei einer zweikanaligen BSS ist jedes Ausgangssignal eine Rekonstruktion eines Quellensignals. Es gilt

$$\begin{aligned} y_1 &= q_1 s_1 + 0 \cdot s_2 \\ y_2 &= 0 \cdot s_1 + q_2 s_2 \end{aligned} \quad \Rightarrow \quad \mathbf{W}\mathbf{H} = \begin{bmatrix} q_1 & 0 \\ 0 & q_2 \end{bmatrix}.$$

Ebenso ist die permutierte Lösung möglich:

$$\mathbf{W}\mathbf{H} = \begin{bmatrix} 0 & q_1 \\ q_2 & 0 \end{bmatrix}.$$

Macht man nun einen Koeffizientenvergleich, folgt, daß die Matrix \mathbf{W} einer der Matrizen $[q_1 \mathbf{w}_{MV_1}^H \quad q_2 \mathbf{w}_{MV_2}^H]$ oder $[q_2 \mathbf{w}_{MV_2}^H \quad q_1 \mathbf{w}_{MV_1}^H]$ entspricht. q.e.d.

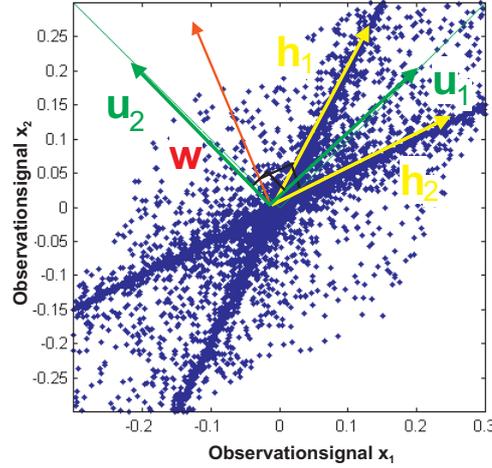


Abb. 6.1: Streudiagramm der Observationssignale mit den Hauptachsen des Konzentrationsellipsoid (Eigenvektoren der Kovarianzmatrix) $\mathbf{u}_1/\mathbf{u}_2$ und dem Steeringvektor des adaptiven Beamformers \mathbf{w}

Lemma 5 kann für weitere Untersuchungen benutzt werden. Die Eigenwertzerlegung von \mathbf{R}_x ist wieder

$$\mathbf{R}_x = \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{U}^H,$$

damit gilt

$$\mathbf{R}_x^{-1} = (\mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{U}^H)^{-1} = \mathbf{U} \cdot \mathbf{\Lambda}^{-1} \cdot \mathbf{U}^H,$$

und für die Bestimmungsgleichung des MV Beamformers (4.16) ergibt sich, wenn die Normalisierung zunächst vernachlässigt wird,

$$\mathbf{w}_{\text{MV}_1} = \mathbf{R}_x^{-1} \hat{\mathbf{h}}_1 = (\mathbf{U} \cdot \mathbf{\Lambda}^{-1} \cdot \mathbf{U}^H) \cdot \hat{\mathbf{h}}_1 = \mathbf{U} \cdot \mathbf{\Lambda}^{-\frac{1}{2}} \cdot \mathbf{\Lambda}^{-\frac{1}{2}} \cdot \mathbf{U}^H \cdot \hat{\mathbf{h}}_1. \quad (6.1)$$

Definiert man nun $\hat{\mathbf{t}}_1$ als

$$\hat{\mathbf{t}}_1 = \mathbf{\Lambda}^{-\frac{1}{2}} \cdot \mathbf{U}^H \cdot \hat{\mathbf{h}}_1 = \begin{bmatrix} \frac{\mathbf{u}_1}{\lambda_1} \\ \frac{\mathbf{u}_2}{\lambda_2} \end{bmatrix} \cdot \hat{\mathbf{h}}_1, \quad (6.2)$$

sind $\hat{t}_{11}/\hat{t}_{12}$ die Koordinaten von $\hat{\mathbf{h}}_1$ im Raum, der vom Orthogonalsystem $\frac{\mathbf{u}_1}{\lambda_1}$ und $\frac{\mathbf{u}_2}{\lambda_2}$ aufgespannt wird. $\hat{\mathbf{t}}_1$ ist also der Verzögerungsvektor, der in den Raum der geweißten Signale transformiert wurde (vgl. auch Abb. 6.1 und Abb. 6.2). Für die Rekonstruktion von s_1 gilt damit

$$\begin{aligned} \hat{s}_1 = \mathbf{w}_{\text{MV}_1}^H \mathbf{x} &= (\mathbf{U} \cdot \mathbf{\Lambda}^{-\frac{1}{2}} \cdot \hat{\mathbf{t}}_1)^H \mathbf{x} \\ &= \hat{\mathbf{t}}_1^H \cdot \mathbf{\Lambda}^{-\frac{1}{2}} \cdot \mathbf{U}^H \cdot \mathbf{x} \\ &= \hat{\mathbf{t}}_1^H \cdot \mathbf{V} \cdot \mathbf{x} = \hat{\mathbf{t}}_1^H \mathbf{z}. \end{aligned} \quad (6.3)$$

Da die geweißten Stör- und Nutzsignale senkrecht aufeinander stehen, ist es plausibel, den in diesem Raum transformierten Verzögerungsvektor des Nutzsignals als Schätzvektor zu verwenden. Er steht damit, wie in Abb. 6.2 sichtbar, senkrecht auf dem Störsignal und minimiert es im Sinne eines MMSE-Schätzers.

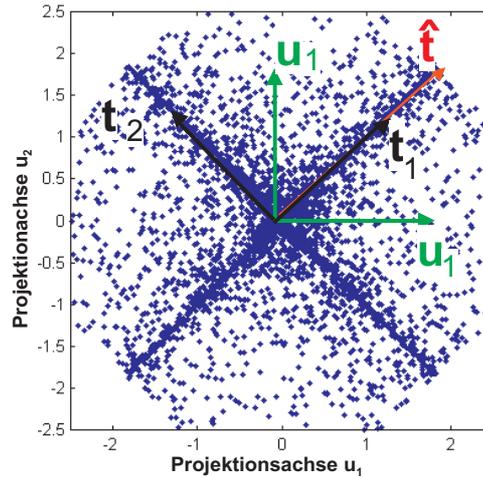


Abb. 6.2: Streudiagramm der geweißten Signale mit den Vektoren der Entmischungsmatrix $\mathbf{t}_1/\mathbf{t}_2$ und dem transformierten Verzögerungsvektor $\hat{\mathbf{t}}_1$

Abb. 6.1 zeigt das Streudiagramm der gemischten Signale¹ mit den Mischungsvektoren und den Eigenvektoren der Kovarianzmatrix (\mathbf{u}_1 , der die Richtung der größten Streuung angibt und \mathbf{u}_2 , der senkrecht dazu steht), den Mischungsvektoren \mathbf{h}_1 und \mathbf{h}_2 sowie den Steeringvektor \mathbf{w}_{MV_1} . In Abb. 6.2 wurden die Signale geweißt. Die Koordinaten sind jetzt in den Richtungen von \mathbf{u}_1 und \mathbf{u}_2 gelegt. \mathbf{t}_1 und \mathbf{t}_2 sind die Mischungsvektoren im Raum der geweißten Signale.

Aus der Äquivalenz der beiden Konzepte folgt, daß $\hat{\mathbf{t}}_1$ und einer der beiden Vektoren der Rotationsmatrix \mathbf{T} in die gleiche Richtung zeigen müssen. Die Längen müssen nicht gleich sein, weil bei der Herleitung von (6.2) die Normierung nicht berücksichtigt wurde und die ICA ohnehin nicht eindeutig gegenüber der Skalierung ist. Damit muß die ICA gegen $\hat{\mathbf{t}}_1$ konvergieren, wenn die obige Voraussetzung eingehalten wurde, daß der Verzögerungsvektor exakt geschätzt worden ist.

6.2 ICA mit geometrischen Randbedingungen

6.2.1 Ableitung des Algorithmus

Die Äquivalenz zwischen adaptiven Beamformern und ICA-basierten blinden Verfahren kann genutzt werden, um die Nachteile der beiden Methoden zu verringern. Ein solches Verfahren wurde von Knaak *et al.* (2003) vorgeschlagen und soll im folgenden beschrieben werden.

Dazu seien N Kanäle bekannt, die durch die Aufnahme von einem Nutzsignal und weiteren Störsignalen entstanden sind. Der zum Signal s_1 gehörige Verzögerungsvektor $\hat{\mathbf{h}}_1$ sei ebenfalls bekannt. Allerdings wird davon ausgegangen, daß er nur grob geschätzt werden kann. Die Gründe für eine ungenaue Schätzung liegen in den in Kap. 4.4.1 besprochenen Problemen (Nachhall und falsche Einfallsrichtung).

¹Wie in den vorhergehenden Fällen wurden nur reellwertige Zufallsvektoren gezeichnet.

Wie in den vorangegangenen Kapiteln wurden die Observationssignale \mathbf{x}^b zunächst mit der STFT in schmalbandige Signalkomponenten \mathbf{x} zerlegt und diese einzeln mit der PCA geweißt $\mathbf{z} = \mathbf{V}\mathbf{x}$.

Nun wird anstelle der Energiefunktion des MV Beamformers (4.14) die in Kap. 5.2.1 eingeführte Energiefunktion des FastICA-Algorithmus verwendet

$$J = (k_2 E\{g_2(|\mathbf{t}_1 \mathbf{z}|^2)\} - E\{g_2(y_{\text{Gauss}})\})^2. \quad (6.4)$$

Die geometrische Randbedingung des ABF wird weiter genutzt: Die Signale aus der Richtung $\hat{\mathbf{h}}_1$ sollen ungehindert durchgelassen werden. Damit ergibt sich nach dem Weißen:

$$\mathbf{w}^H \hat{\mathbf{h}}_1 = c_1 = \mathbf{t}_1^H \mathbf{V} \hat{\mathbf{h}}_1. \quad (6.5)$$

Da eine geometrische Randbedingung genutzt wird, wird dieser Algorithmus im folgenden „Geometrically Constrained ICA“ (CICA) genannt.

Bei einem FastICA-Algorithmus wird auch (6.4) optimiert, jedoch mit der Randbedingung, daß $\hat{\mathbf{t}}_1$ auf dem Einheitskreis bleibt. Die Randbedingung wird damit begründet, daß die Entmischungsmatrix der geweißten Signale eine unitäre Matrix sein muß (vgl. Kap. 5.1.4) und ihre Vektoren folglich die Länge eins haben müssen.

Die Bedingung kollidiert mit der geometrischen Randbedingung, wie in Abb. 6.3 deutlich wird. Diese Abbildung zeigt das Streudiagramm der geweißten Signale mit der in den Raum der geweißten Signale transformierten geometrischen Randbedingung (Gerade) sowie der Randbedingung der FastICA (Kreis). Da die Gerade den Kreis im allgemeinen Fall nicht berührt, können beide Bedingungen nicht gleichzeitig erfüllt werden.

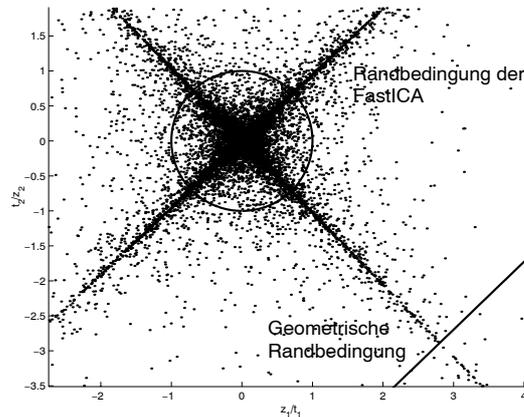


Abb. 6.3: Streudiagramme der geweißten Signale mit den Randbedingungen von CICA und FastICA

In Kap. 5.2.1 werden die Kuhn-Tucker-Punkte der FastICA mit

$$E\{g_2'(|\mathbf{t}_1^H \mathbf{z}|^2) \cdot \mathbf{z} \cdot (\mathbf{t}_1^H \mathbf{z})^*\} = \alpha \mathbf{t}_1 \quad (6.6)$$

angegeben. Diese Punkte können auch als stationäre Punkte für das Optimierungsproblem mit der geometrischen Randbedingung benutzt werden. Ist \mathbf{t}_1 ein Kuhn-Tucker-Punkt auf dem Einheitskreis, dann ist $\mathbf{t}'_1 = \beta \mathbf{t}_1$ für ein beliebiges β ein Kuhn-Tucker-Punkt derjenigen Lagrange-Funktion, die den Raum der Lösungen auf Vektoren der Länge β begrenzt. Entsprechend der Informationstheorie ist die Negentropie invariant gegenüber der Skalierung. Da die Länge des Entmischungsvektors nur die Amplitude des geschätzten Signals ändert, ist es ausreichend (und geboten), nur Vektoren gleicher Länge zu vergleichen. Das heißt für das besprochene Problem, daß man nach einem Vektor sucht, der die Randbedingung erfüllt und gleichzeitig von allen Vektoren gleicher Länge die maximale Entropie hat.

Damit ändert sich die Lösung nicht, wenn man sie auf die geometrische Randbedingung projiziert. Der resultierende Algorithmus besteht aus zwei Schritten: Der Berechnung eines neuen Wertes mit FastICA (6.7) und Projektion auf die Randbedingung (6.8). Abb. 6.4 illustriert die Iteration.

$$\mathbf{t}_{1_{k+1}} = \mathbf{t}_{1_k} - \frac{\mathbb{E}\{g'_2(|\mathbf{t}_{1_k}^H \mathbf{z}|^2) \cdot \mathbf{z} \cdot (\mathbf{t}_{1_k}^H \mathbf{z})^*\} - \alpha \mathbf{t}_{1_k}}{\mathbb{E}\{g'_2(|\mathbf{t}_{1_k}^H \mathbf{z}|^2)\} + \alpha} \quad (6.7)$$

$$\mathbf{t}_{1_{k+1new}} = \frac{\mathbf{t}_{1_{k+1}}}{|\mathbf{t}_{k+1}^H \mathbf{V} \hat{\mathbf{h}}_1|} \quad (6.8)$$

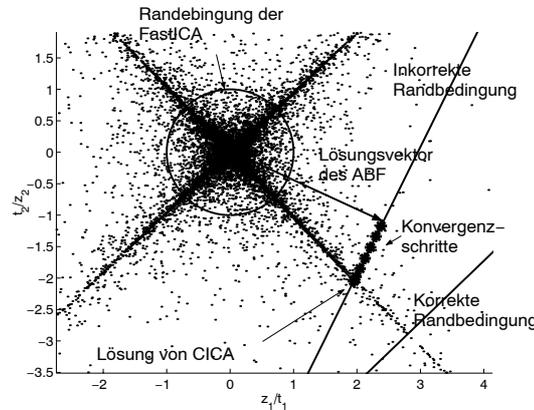
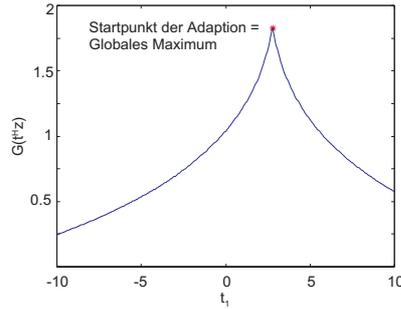
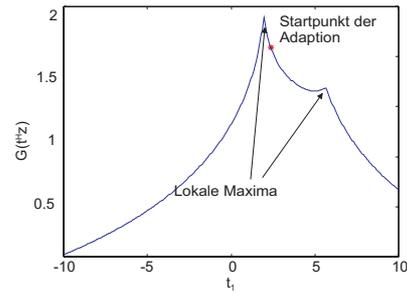


Abb. 6.4: Konvergenz der CICA: Ausschnitt aus dem Streudiagramm der geweißten Signale mit dem Startpunkt der Iteration sowie der Randbedingung. Die Punkte geben die Konvergenz zur richtigen Lösung an.

Der neue Algorithmus startet mit $\mathbf{t}_{1_0} = \mathbf{V} \hat{\mathbf{h}}_1$. Wenn die Schätzung von $\hat{\mathbf{h}}_1$ korrekt ist, ist das bereits, wie in Kap. 6.1 nachgewiesen, die richtige Lösung, so daß keine weitere Iteration nötig ist (Abb. 6.5a). In anderen Fällen konvergiert der Algorithmus durch wiederholtes Anwenden von (6.7) und (6.8) zu den Punkten maximaler Negentropie (Abb. 6.5b).



(a) Korrekte Randbedingung, $\mu = 0^\circ$,
 $\alpha = 27^\circ$



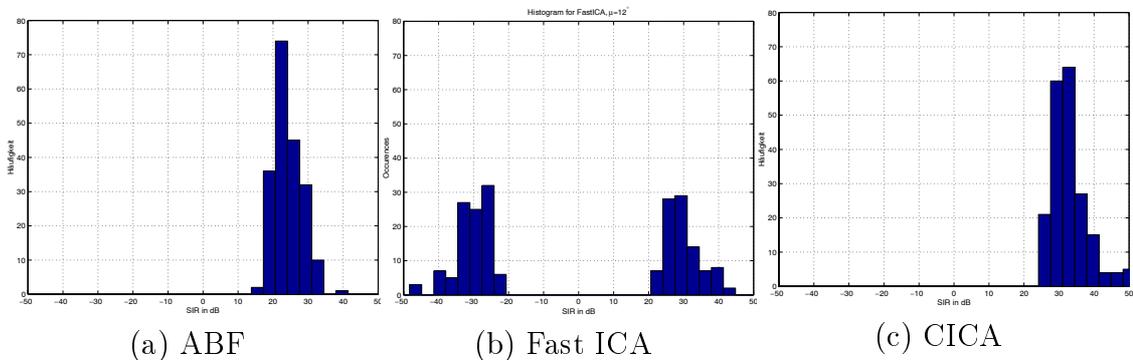
(b) Fehlerhafte Randbedingung,
 $\mu = -12^\circ$, $\alpha = 27^\circ$

Abb. 6.5: Energiefunktion auf einer korrekt und fehlerhaft geschätzten Randbedingung mit den Startpunkten

6.2.2 Simulationen

Zum Test des Algorithmus wurden 200 Überlagerungen simuliert. Dazu wurden zwei komplexe Zufallssignale gleicher Varianz mit zufällig gewählten, komplexen Filtermatrizen gemischt. Das Nutzsignal (das jeweils erste Zufallssignal) wurde nun mit einem ABF (MV Beamformer), der CICA und der FastICA geschätzt. Bei letzterer wurde das erste Ausgangssignal benutzt (keine Lösung des Permutationsproblems). Für die beiden ersten wurde der zum ersten Signal gehörige Mischungsvektor verwendet und für die Tests mit einem fehlerhaften Verzögerungsvektor zufällig so modifiziert, daß ein Schätzfehlerwinkel von $\mu = 12^\circ$ entstand (vgl. Kap. 4.4.1). Nach jeder Simulation wurde die Störunterdrückung bestimmt, die Störunterdrückungen aller Tests sind in den Histogrammen der folgenden Abbildungen dargestellt.

Wenn der richtige Verzögerungsvektor verwendet wurde, erreichen der ABF und die CICA eine sehr gute Störunterdrückung (Abb. 6.6a und Abb. 6.6c). In keinem Fall war der SIR kleiner als 10 dB. FastICA zeigt zwei Cluster: einen mit einer guten Rauschunterdrückung und einen mit einem negativen SIR (Abb. 6.6b). Die Ursache liegt im ungelösten Permutationsproblem: Teilweise schätzt der Algorithmus das Störsignal und unterdrückt das Nutzsignal.



(a) ABF

(b) Fast ICA

(c) CICA

Abb. 6.6: Histogramme für die Störunterdrückung mit einem korrekten Verzögerungsvektor (Simulationen)

Wird der Verzögerungsvektor nur grob geschätzt, kann mit einem ABF keine Störunterdrückung erreicht werden (Abb. 6.7a). Werte um 0 dB, d.h. einen jeweils gleich lauten Anteil des Nutz- und des Störgeräuschs, sind hier die Regel. Die Ursachen für diese schlechte Ergebnis wurden in Kap. 4.4.3 ausführlich diskutiert. Da die FastICA den Verzögerungsvektor nicht verwendet, gilt das Resultat aus Abb. 6.6b auch hier. Die CICA schätzt in den meisten Fällen das Nutzsignal richtig und unterdrückt das Störsignal sehr gut (Abb. 6.7b). Nur in seltenen Fällen konvergiert sie gegen das Störsignal. Darum wird im folgenden eine analytische Bewertung des zugelassenen Fehlers bei der Schätzung des Verzögerungsvektors vorgenommen.

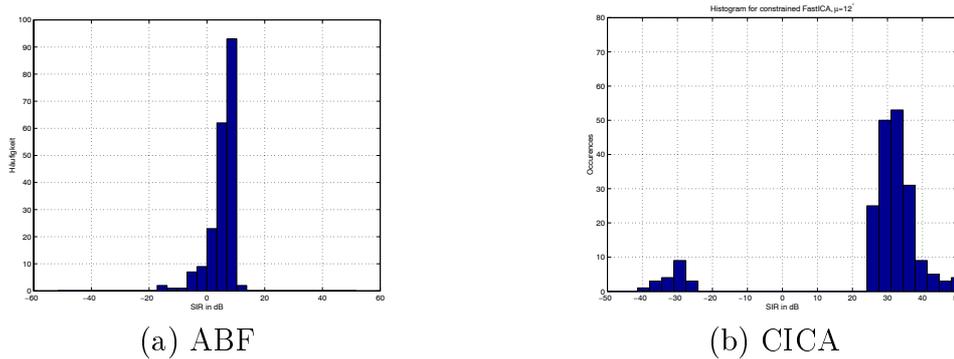


Abb. 6.7: Histogramme für die Störunterdrückung mit einem fehlerhaften Verzögerungsvektor $\mu = 12^\circ$ (Simulationen)

Die Abb. 6.8 und Abb. 6.9 zeigen die gleichen Resultate für $N = M = 3$, d.h. das simulierte Nutzsignal wurde mit zwei Störquellen überlagert, und drei Observationen wurden ausgewertet. Das Ergebnis deckt sich mit dem zweikanaligen Fall.

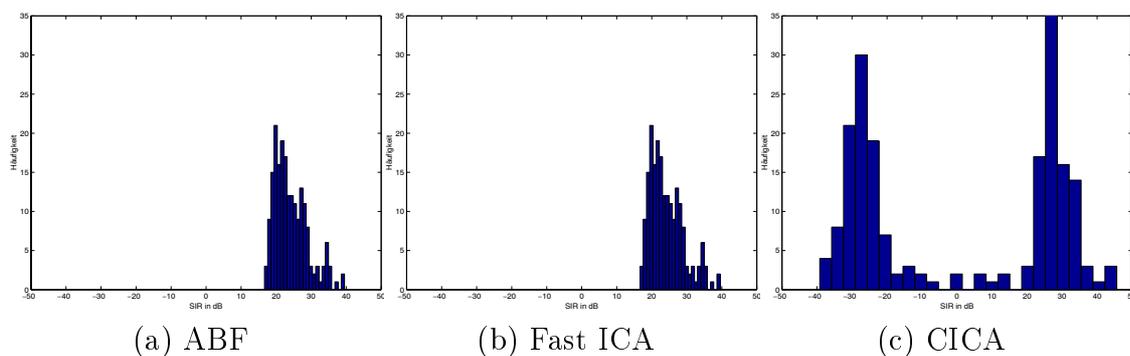


Abb. 6.8: Histogramme für die Störunterdrückung für $N = M = 3$ mit einem korrekten Verzögerungsvektor (Simulationen)

6.2.3 Analytische Untersuchung der Konvergenz

Um die analytische Betrachtung zu vereinfachen und um auf die in Kap. 4.4 gewonnenen Erkenntnisse zurückgreifen zu können, wird die Betrachtung auf $N = M = 2$

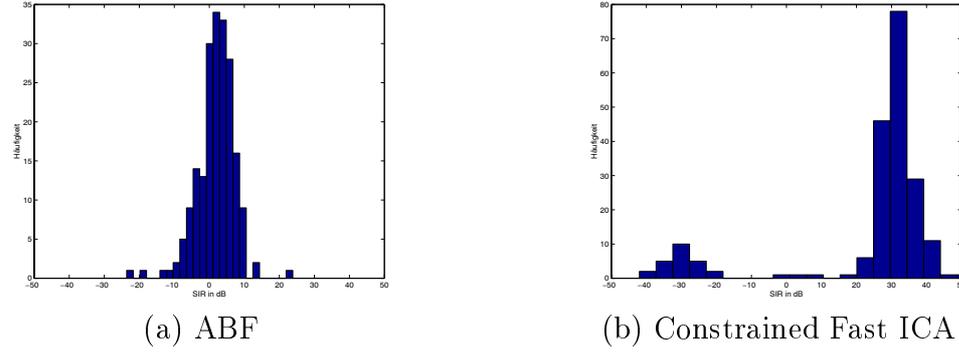


Abb. 6.9: Histogramme für die Störunterdrückung für $N = M = 3$ mit einem fehlerhaften Verzögerungsvektor $\mu = 12^\circ$ (Simulationen)

beschränkt. Die Resultate konnten jedoch an praktischen Beispielen auch für den höherdimensionalen Fall bestätigt werden.

Wird die Nebenbedingung richtig geschätzt, existiert, wie in Abb. 6.5 zu sehen, nur ein Maximum der Lagrangefunktion. Bei einer ungenauen Schätzung hingegen existieren zwei lokale Maxima, die zu den beiden Signalen gehören. Der Algorithmus konvergiert zu einem der beiden Maxima, was mit der Simulation übereinstimmt. Ob

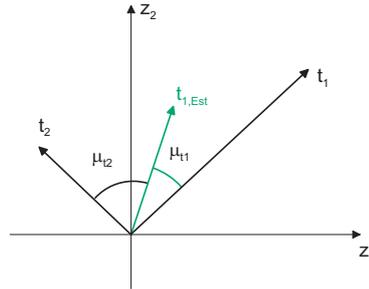


Abb. 6.10: Vektoren im Raum der geweißten Signale

er zum globalen Maximum konvergiert, wird von der Form der Lagrange-Funktion und vom Startpunkt festgelegt. Beide Parameter sind Funktionen des schon in Kap. 4.4.1 eingeführten Schätzfehlerwinkels (μ), der die Schätzgenauigkeit beschreibt. Seine Transformation in den Raum der geweißten Signale ist μ_{t_1} , d.h. der Winkel zwischen $\hat{\mathbf{t}}_1$ und $\mathbf{t}_{1,opt} = \mathbf{V}\mathbf{h}_1$ (vgl. Abb. 6.10). Benutzt man die Resultate aus Kap. 4.4.3, ergibt sich, wie folgt:

$$\begin{aligned}
 \cos(\mu_{t_1}) &= \frac{\hat{\mathbf{t}}_1^H \mathbf{t}_1}{|\hat{\mathbf{t}}_1| |\mathbf{t}_1|} = \frac{\hat{\mathbf{h}}_1^H \mathbf{V}^H \mathbf{V} \mathbf{h}_1}{|\hat{\mathbf{t}}_1| |\mathbf{t}_1|} = \frac{\hat{\mathbf{h}}_1^H \mathbf{U} \mathbf{\Lambda}^{-1/2} \mathbf{\Lambda}^{-1/2} \mathbf{U}^H \mathbf{h}_1}{|\hat{\mathbf{t}}_1| |\mathbf{t}_1|} \\
 &= \frac{\hat{\mathbf{h}}_1^H \mathbf{U} \mathbf{\Lambda}^{-1} \mathbf{U}^H \mathbf{h}_1}{|\hat{\mathbf{t}}_1| |\mathbf{t}_1|} \frac{\hat{\mathbf{h}}_1^H \mathbf{R}_x^{-1} \mathbf{h}_1}{|\hat{\mathbf{t}}_1| |\mathbf{t}_1|} = \frac{\cos \mu - \cot \alpha \sin \mu}{|\hat{\mathbf{t}}_1| |\mathbf{t}_1|}.
 \end{aligned} \tag{6.9}$$

Analog läßt sich der Winkel μ_{t_2} zwischen $\hat{\mathbf{t}}_1$ und dem transformierten Verzögerungs-

vektor $\hat{\mathbf{t}}_{2opt} = \mathbf{V}\mathbf{h}_2$ definieren:

$$\cos(\mu_{t_2}) = \frac{\hat{\mathbf{t}}_1^H \mathbf{t}_1}{|\hat{\mathbf{t}}_1| |\mathbf{t}_1|} = \frac{\frac{\sin \mu}{\sin \alpha}}{|\hat{\mathbf{t}}_1| |\mathbf{t}_2|}.$$

Die Konvergenz zum Nutzsignale ist garantiert, solange $\cos(\mu_{t_2}) > \cos(\mu_{t_1})$ ist (vgl. Abb. 6.10).

$$\begin{aligned} \cos(\mu_{t_1}) &< \cos(\mu_{t_2}) \\ \frac{\cos \mu - \cot \alpha \sin \mu}{|\hat{\mathbf{t}}_1| |\mathbf{t}_1|} &< \frac{\frac{\sin \mu}{\sin \alpha}}{|\hat{\mathbf{t}}_1| |\mathbf{t}_2|} \\ \cos \mu - \cot \alpha \sin \mu &< \frac{\sin \mu}{\sin \alpha} \\ \cos \mu &< \cot \alpha \sin \mu + \frac{\sin \mu}{\sin \alpha} \\ \cot \mu &< \frac{1 + \cos \alpha}{\sin \alpha} \end{aligned} \quad (6.10)$$

$$\cot \mu_{\text{critical}} = \frac{1 + \cos \alpha}{\sin \alpha} \Rightarrow \mu_{\text{critical}} = \frac{\alpha}{2} \quad (6.11)$$

Der Winkel α wird in in (4.20) und in Abb. 4.6b als Winkel zwischen den Verzögerungsvektoren des Nutz- und des Störsignals in der Streuebene definiert.

Aus (6.10) kann man eine hinreichende Bedingung für die Konvergenz der CICA ableiten: Der Algorithmus konvergiert dann gegen das lokale Minimum, das zum Nutzsignal gehört, wenn $\mu < \mu_{\text{critical}}$ ist. Den kritischen Winkel von $\alpha/2$ kann man sich im hallarmen Raum so veranschaulichen, daß der geschätzte Verzögerungsvektor eher in die Richtung des Nutzsignals als in die Richtung des Störsignals zeigt.

Zur Illustration dienen die Streudiagramme in Abb. 6.11. Sie zeigen die Randbedingungen für drei verschiedene Fälle. Im ersten Fall (Abb. 6.11a/b) stimmt der Verzögerungsvektor exakt, so daß sowohl die CICA als auch der MV-Beamformer den richtigen Entmischungsvektor für das Nutzsignal finden. Im zweiten Fall (Abb. 6.11c/d) ist der Verzögerungsvektor nicht richtig geschätzt, der Fehler ist jedoch kleiner als der kritische Winkel, denn im Raum der geweißten Signale in Abb. 6.11d erkennt man, daß hier offenbar $\cos \mu_{t_1} > \cos \mu_{t_2}$ gilt. Nur die CICA findet die richtige Lösung in diesem Fall, während der Steeringvektor des MV-Beamformers nicht senkrecht auf dem Störsignal steht. Im dritten Fall (Abb. 6.11e/f) ist $\mu > \mu_{\text{critical}}$, und die CICA konvergiert gegen die permutierte Lösung. Anschaulich bedeutet dies, daß der geschätzte Verzögerungsvektor näher an \mathbf{h}_2 als an \mathbf{h}_1 liegt. Der Algorithmus fokussiert sich auf das Störsignal.

Die Bedingung in (6.10) erlaubt einen Vergleich zwischen dem MV-Beamformer und der CICA: Die CICA ist wesentlich robuster gegenüber einer Fehlschätzung des Verzögerungsvektors als ein MV-Beamformer. Die SIRs beider Methoden sind in Abhängigkeit vom Schätzfehlerwinkel in Abb. 6.12 aufgezeichnet. Der SIR des ABF sinkt, sobald der Verzögerungsvektor nicht ganz exakt bestimmt ist. Die CICA schätzt das Signal optimal, solange μ kleiner als μ_{critical} ist. Danach fokussiert er sich auf das Störsignal, was in einem hohen, negativen SIR resultiert.

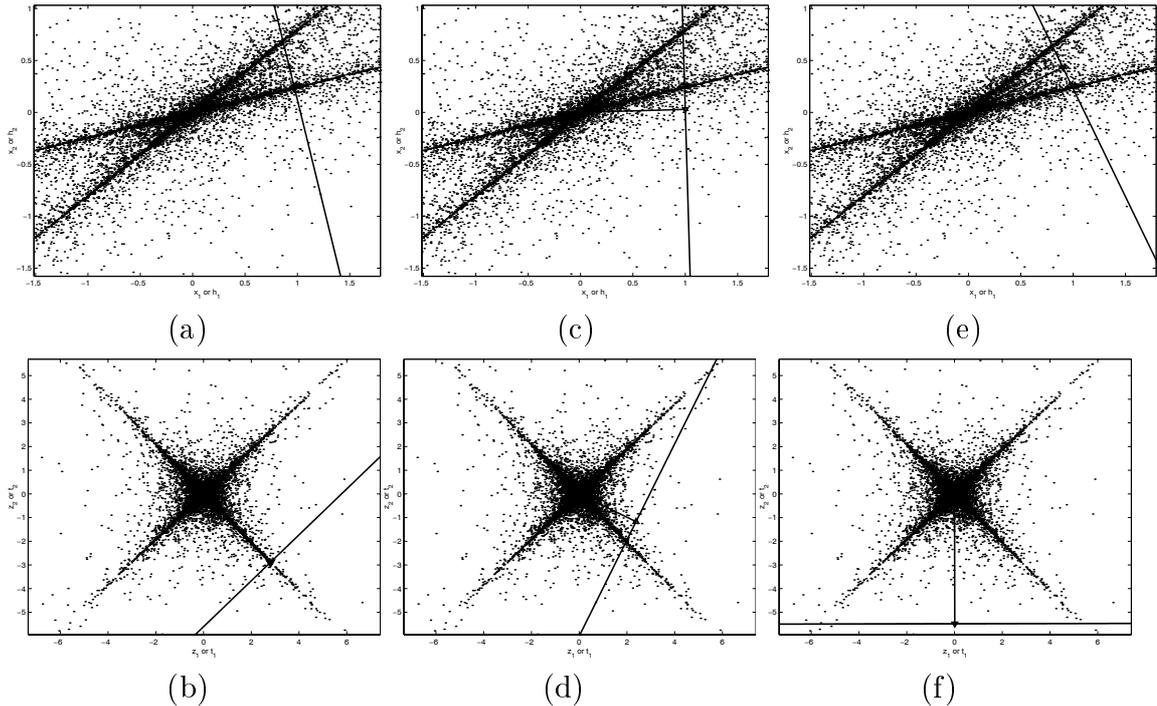


Abb. 6.11: Streudiagramme für \mathbf{x} und \mathbf{z} mit verschiedenen Randbedingungen, die sich aus unterschiedlich genau geschätzten Verzögerungsvektoren ergeben, und den Schätzvektoren des MV-Beamformers und der CICA

Für den praktischen Einsatz stellt sich die Frage, wann die Bedingung (6.10) eingehalten wird. Nach Kap 4.4.1 sind zwei Gründe für eine falsche Schätzung verantwortlich: eine ungenaue Einfallsrichtung und der Nachhall. Da μ nach (6.10) nur kleiner als $\alpha/2$ sein muß, muß von der Schätzung der Einfallsrichtung lediglich gefordert werden, daß der geschätzte Verzögerungsvektor näher an der Nutz- als an der Störquelle liegt². Die CICA fokussiert sich also (im hallfreien Raum) auf die dem geschätzten Verzögerungsvektor am nächsten gelegene Quelle. Dies geschieht unabhängig von der Qualität der Schätzung. Der Algorithmus ist also sehr robust.

Bei einer Verfälschung des Verzögerungsvektors durch nicht berücksichtigten Nachhall kann es zu größeren Abweichungen kommen. In jedem Frequenzbereich wird $\hat{\mathbf{h}}_1$ nur über die Phasendrehung infolge der Verzögerung bestimmt, während in \mathbf{h}_1 alle Reflexionen einfließen. Der dadurch entstehende Fehler sowie der kritische Winkel sind für einen realen Raum mit einer Nachhallzeit von 150 ms (vgl. Kap. 8.1) in Abb. 6.13 dargestellt. Es zeigt sich, daß μ auch in einem Raum mit Nachhall, abgesehen von singulären Punkten, immer unterhalb des kritischen Winkels bleibt.

Lediglich für sehr kleine Frequenzen kann der Algorithmus nicht eingesetzt werden. Die Phasendifferenzen sind hier sehr gering und damit α und der kritische Winkel ebenfalls sehr klein, so daß es vermehrt zu falschen Permutationen kommt. In diesem Bereich sollte die klassische FastICA mit dem Ansatz von Murata *et al.*

²Eine richtige Lösung für den Fall, daß der Verzögerungsvektor auf die Störquelle zeigt, würde darauf hindeuten, daß er entweder irrelevant ist oder der Algorithmus „hellseherisch“ arbeitet.

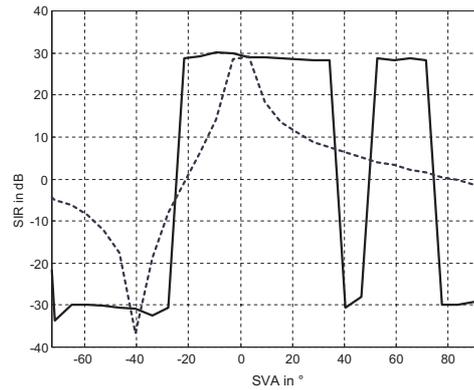
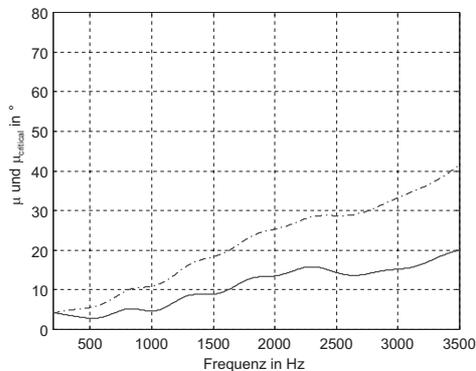
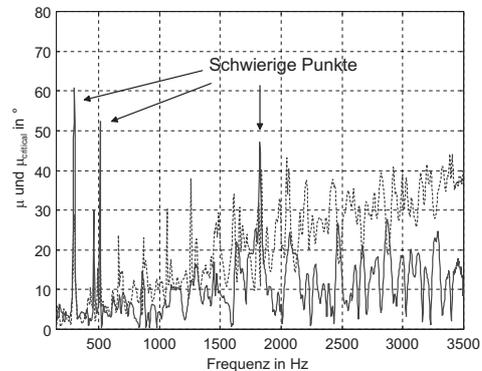


Abb. 6.12: Vergleich der Robustheit der CICA (durchgezogene Linie) und des MV-Beamformers (gestrichelte Linie)



(a) Kein Nachhall ($t_r \leq 150 \text{ ms}$)



(a) Nachhallzeit $t_r = 150 \text{ ms}$

Abb. 6.13: Vergleich realer Schätzfehler (durchgezogen) mit den kritischen Winkeln (gestrichelt) für reale Mischungen

(1999) zur Permutationsvermeidung verwendet werden. Werden diese Frequenzabschnitte am Ende berechnet, kann die Amplitudenmodulation der schon mit der CICA berechneten Frequenzabschnitte als Referenz genutzt werden. Anhand der Übereinstimmung zu diesen gemäß (5.53) wird das Permutationsproblem einfach gelöst.

Kapitel 7

Anwendung auf Maschinensignale

In diesem Kapitel sollen die in den Kap. 4 - Kap. 6 vorgestellten bzw. hergeleiteten Algorithmen auf die Maschinendiagnose angewendet werden. Dazu muß zunächst geprüft werden, ob die bei der Herleitung gemachten Annahmen auch für Motorsignale gelten, die mit den üblichen Geräuschen einer Fertigungshalle gestört sind. Die wichtigsten Annahmen sind die statistische Unabhängigkeit bzw. die Unkorreliertheit der Geräusche untereinander. Die Plausibilität dieser Annahmen wird anhand realer Daten aus der Fertigung von Kleinmotoren bestätigt.

Weiterhin kann man die Konvergenz der ICA-Algorithmen verbessern, wenn man - entgegen der eigentlich angenommenen „Blindheit“ - Vorwissen, wie die Wahrscheinlichkeitsverteilungen der zu rekonstruierenden Signale, nutzt und die Parameter, wie die Nichtlinearität der ICA, darauf abstimmt. Die Prüfung der Voraussetzungen und das „Feintuning“ erfolgt in Kap. 7.1.

In der Literatur findet man nur wenige Anwendungen der ICA auf die technische Diagnose. Lediglich die Arbeiten von Servière und Fabry (2001), Fabry und Servière (2000) sowie Ypma und Leshem (2000) verwenden die ICA zur Trennung von Körperschallsignalen. Ypma und Leshem (2000) gehen von skalaren Mischungen der Körperschallsignale aus, während Servière und Fabry (2001) diese Mischungen auch als konvolutiv modellieren. Ihre Algorithmen verwenden im wesentlichen die instationäre Dekorrelation im Frequenzbereich (vgl. Kap. 5.3) mit zusätzlichen Maßnahmen gegen additives unkorreliertes Rauschen. Genaue Untersuchungen über die Gültigkeit der Voraussetzungen wurden bisher noch nicht vorgenommen.

Letztlich bietet die Konzentration auf die konkrete Anwendung die Chance, optimierte Spezialverfahren zu entwickeln. Als „Preis“ für den Verzicht auf die Allgemeingültigkeit kann man damit oft Algorithmen erhalten, die bei geringerer Komplexität - und damit geringeren Kosten oder höherer Robustheit - zu ähnlich guten Ergebnissen führen. Das kann unter Berücksichtigung ökonomischer Rahmenbedingungen ein entscheidender Vorteil für einen praktischen Einsatz sein.

Ein solcher Spezialfall liegt vor, wenn die Störgeräusche nicht kontinuierlich auftreten, während das Nutzsignal, d.h. das Signal einer rotierenden Maschine, kontinuierlich ist. Dann reicht es, aus die Nebengeräuschpausen zu detektieren und in diesen das Nutzsignal zu klassifizieren. Die Methode ist auch dann einsetzbar, wenn in den

„Nebengeräuschpausen“ erheblicher diffuser Schall vorhanden ist, welcher aber die Ausprägung vieler Merkmale nicht beeinflusst. Ein Algorithmus, der die Anzahl der Quellen aus einer mehrkanaligen Aufnahme schätzt, wird in Kap. 7.2 vorgestellt.

7.1 Voraussetzung für die stochastischen Algorithmen

Um die in den vorderen Kapiteln zur mehrkanaligen Signalverarbeitung vorgestellten Algorithmen als Vorverarbeitung für die Klassifikation anwenden zu können, muß zunächst die Gültigkeit der Voraussetzungen überprüft werden, unter denen diese abgeleitet wurden. Die Beamformer (Kap. 4) fordern, daß die Maschinensignale s_{dut}^b und die industriellen Nebengeräusche s_s^b unkorreliert sind. Für die ICA-Verfahren (Kap. 5 und Kap. 6) ist sogar eine statistische Unabhängigkeit nötig. Die Plausibilität dieser Voraussetzungen soll im folgenden für real aufgenommene Daten gezeigt werden.

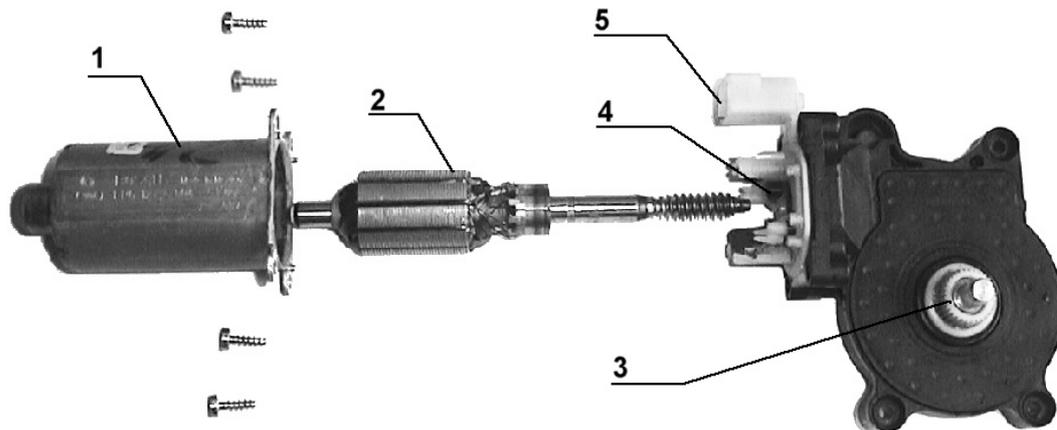


Abb. 7.1: Abbildung des untersuchten Motors nach Kunter (2002)

Dazu werden die akustischen Signale von elektrischen Kleinmotoren (Gleichstrommaschinen für Fensterheber im PKW) in einem schalltoten Raum aufgenommen. Abb. 7.1 zeigt einen solchen Motor mit dem Stator (1), dem Rotor (2) und dem Planetengetriebe (3) zur Erzeugung des erforderlichen Antriebsmoments. Die elektrischen Anschlüsse und die Kohlebürsten sind mit (4) und (5) bezeichnet. Weiterhin wurden Geräuschproben aus der Fertigungshalle dieser Motoren¹ ungefähr am Ort der derzeitigen manuellen akustischen Güteprüfung aufgenommen. Die Proben beinhalten das Klicken von Endabschaltern, Geräusche mechanischer Bearbeitung wie Sägen, Bohren oder Fräsen, den Klang anderer rotierender Maschinen (Lüfter, Antriebe der Werkzeugmaschinen etc.) sowie Rufe von Mitarbeitern (d.h. menschliche Sprache).

Da die Korrelation in jedem Frequenzabschnitt verschwinden muß, werden die Signale s_{dut}^b und s_s^b mit einer STFT in den Frequenzbereich transformiert. Nun muß

¹In der Bosch Mexico S.A. Toluca.

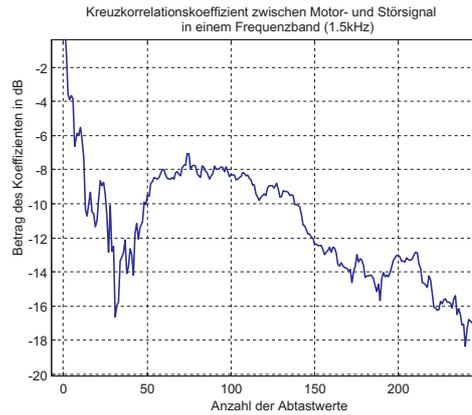


Abb. 7.2: Korrelation zwischen den Störgeräuschen und den Motorsignalen (Kreuzkorrelationskoeffizient)

für jeden Frequenzabschnitt gelten $E\{s_{\text{dut}}^* s_s\} = 0$. Zur Überprüfung werden die Auto- und Kreuzkorrelationskoeffizienten aus dem zeitlichen Mittel geschätzt. Wie man in Abb. 7.2 erkennen kann, verschwindet der Kreuzkorrelationskoeffizient, so daß die Annahme als gerechtfertigt angesehen werden kann.

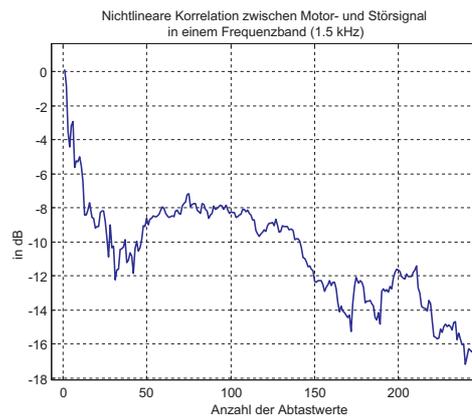


Abb. 7.3: Schätzung der statistischen Abhängigkeit zwischen den Störgeräuschen und den Motorsignalen

Für die ICA-Algorithmen müssen die Störungen statistisch unabhängig zu den Motorsignalen sein, d.h. (5.3) muß gelten: $\forall g : E\{g(s_{\text{dut}}^*) s_s\} = 0$. Zur Überprüfung der Plausibilität wird die Aussage mit dem zuvor beschriebenen Schätzer (zeitliches Mittel) für die nichtlineare Funktion $g = \tanh s$ getestet. Abb. 7.3 zeigt, daß die Annahme plausibel ist. Damit sind die Beamformer und die ICA-Verfahren für die Rekonstruktion von Motorsignalen grundsätzlich anwendbar.

Schließlich wurde in Kap. 5 gezeigt, daß sowohl der Infomax als auch die FastICA und die darauf basierende CICA eine hinreichend gute Annahme für die Wahrscheinlichkeitsdichtefunktion der Quellen benötigen, um eine geeignete Kontrastfunktion g zu wählen. Obwohl eine Konvergenz bereits dann gesichert ist, wenn das Vorzei-

chen der Kurtosis der angenommenen Verteilung mit dem der Kurtosis der Quellen übereinstimmt, führt eine bessere Übereinstimmung auf eine schnellere Konvergenz. Dies soll dazu ausgenutzt werden, die allgemeinen ICA-Algorithmen für Motorsignale zu optimieren. Gemäß der Definition 1 werden die Verteilungen komplexer Zufallsvariablen hier nur auf dem Betrag definiert.

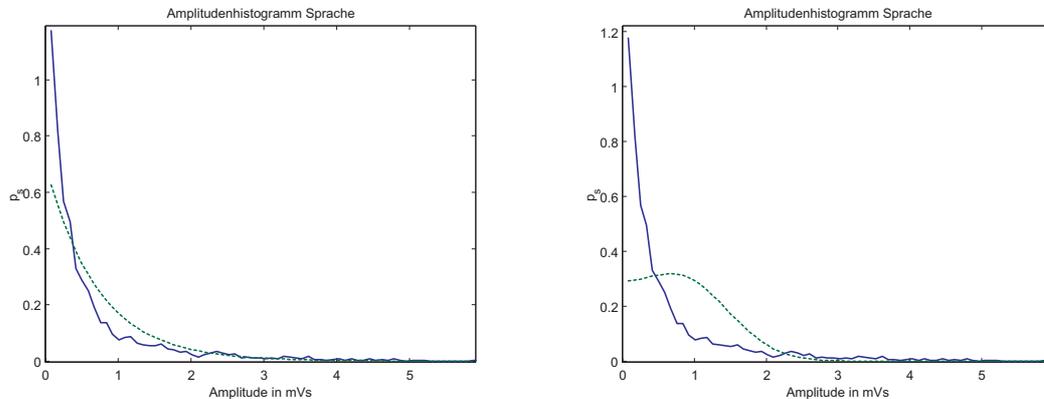


Abb. 7.4: Amplitudenhistogramm einer spektralen Komponente eines Sprachsignals und jeweils eine leptokurtische (a) und eine platykurtische (b) Testverteilung jeweils gleicher Varianz

Nahezu alle Veröffentlichungen² über die Trennung akustischer Mischungen behandeln die Trennung von Sprachsignalen, die eine leptokurtische Verteilung besitzen. Sie kann als eine Laplace-Verteilung modelliert werden, die einen ausgeprägten Peak bei Null besitzt, weil jede Frequenz nur selten angeregt wird, und lange Enden („long tails“) hat und weil bei Knalllauten hohe Signalamplituden entstehen. Abb. 7.4 zeigt das Amplituden-Histogramm der spektralen Komponenten eines Sprachsignals sowie die Laplace-Verteilung gleicher Varianz. Man erkennt eine gute Übereinstimmung, die Sprachsignale sind nur noch „leptokurtischer“³.

Motorsignale haben hingegen eine grundsätzlich andere Verteilung, wie das Histogramm in Abb. 7.5a zeigt. Es stellt das Amplitudenhistogramm der spektralen Komponente bei 1 kHz des Motorsignals dar. Die Form der Wahrscheinlichkeitsdichtefunktion kann man leicht erklären: Die Amplituden der spektralen Komponenten wurden in (5.58) mit

$$|s_{\text{dut}}|^2(t) = a_0^2 + \sum_l a_l \cdot \cos(2\pi l f_0 t + \phi_l)$$

angegeben. Ohne die periodische Amplitudenmodulation wäre die Verteilungsfunktion ein Delta-Impuls bei $p(s_{\text{dut}}) = \delta(s - a_0)$. Durch die Amplitudenmodulation wird dieser verbreitert. Da sie keinen starken Peak bei Null besitzt, kann man sie nicht als Laplace- oder Normalverteilung ansehen und muß sie mit einer platykurtischen Verteilungsfunktion modellieren.

²Lediglich Servière und Fabry (2001) verwenden wie erwähnt Motorsignale, machen aber keine explizite Schätzung der zugrunde liegenden Wahrscheinlichkeitsdichte.

³In der Literatur werden solche Verteilungen oft auch als „peaky“ bezeichnet.

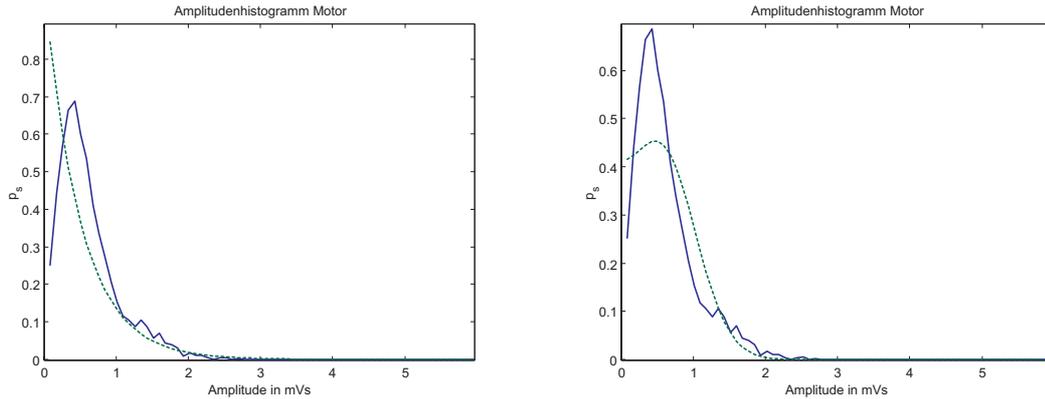


Abb. 7.5: Amplitudenhistogramm einer spektralen Komponente eines Motorsignals mit einer leptokurtischen (a) und platykurtischen Testverteilung (b) $p(s_{\text{dut}}) = b \cosh cs \cdot e^{-s^2/2}$ mit $c = 1, 8$, alle Verteilungen mit gleicher Varianz

Als Modellansatz wird eine modifizierte Normalverteilung benutzt: $p(s_{\text{dut}}) = b \cosh cs \cdot e^{-s^2/2}$, die in Abb. 7.5b für $c = 1, 8$ mit einer durchgezogenen Linie eingezeichnet ist. In der Literatur, z.B. Hyvärinen *et al.* (2001), werden platykurtische Verteilungen häufig mit $c = 1$ benutzt. Diese Verteilung ist eine aufgeweitete Normalverteilung mit einer Häufung bei Null (gestrichelte Linie in Abb. 7.6). Die charakteristische Eigenschaft der Amplitudenverteilung rotierender Maschinen, eine Häufung bei a_0 zu besitzen, kann damit nicht modelliert werden. Darum führt die Verwendung von $c > 1$ zu einer schnelleren und robusteren Konvergenz. Die Kontrastfunktion für g_n (vgl. (5.27)) ergibt sich damit zu

$$\begin{aligned} g_n(y) &= (\log p)' = \frac{p'}{p} = \frac{-2y \cdot b \cosh cy \cdot e^{-y^2} + bc \cdot \sinh cy \cdot e^{-y^2}}{b \cosh cy \cdot e^{-y^2}} \\ &= c \tanh cy - y \end{aligned} \quad (7.1)$$

und die Ableitung von g_n zu

$$g'(x) = c^2 - 1 - c^2 \tanh^2 cy.$$

Die Nebengeräusche sind sehr heterogen: Die Geräusche anderer rotierender Maschinen haben ähnliche Verteilungsfunktionen wie die Testmotoren. Sprachsignale haben, wie gezeigt, eine Laplaceverteilung. Das gilt auch für das „Klicken“ von Schaltern oder Anschlagen von Hubtischen⁴. Andere stellen eine Überlagerung verschiedener Quellen dar und sind nach dem zentralen Grenzwertsatz nahezu normalverteilt. In den Messungen, die nachfolgend für Testzwecke verwendet werden, dominieren platykurtische Störsignale, vgl. Abb. 7.7. Obwohl in der Praxis auch leptokurtische Signale dominieren können, wird im folgenden mit der Kontrastfunktion aus (7.1) gearbeitet, weil die exakte Rekonstruktion der Nebengeräusche nicht das Ziel dieser Arbeit ist.

⁴Diese Signale haben noch häufiger Signalpausen als Sprachsignale, treten dann aber mit großer Amplitude auf. Darum sind sie noch stärker leptokurtisch als die Verteilung eines Sprachsignals.

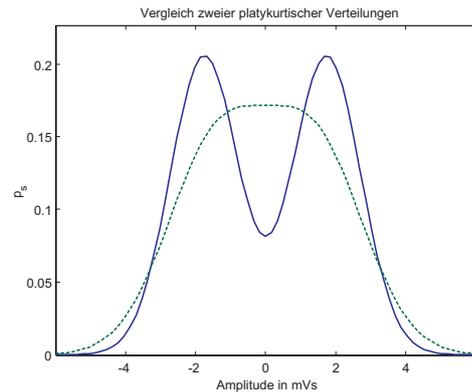


Abb. 7.6: Zwei platykurtische Testverteilungen mit $c = 1,8$ (durchgezogen) und mit $c = 1$ (gestrichelt), gleiche Varianz

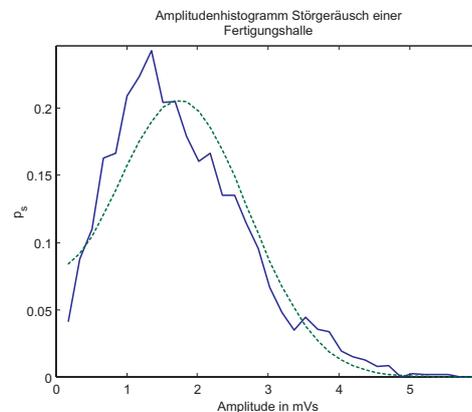


Abb. 7.7: Amplitudenhistogramm einer spektralen Komponente eines Störsignals mit der platykurtischen Testverteilung $p(s_{\text{dut}}) = b \cosh cs \cdot e^{-s^2/2}$ und $c = 2,5$, alle Verteilungen mit gleicher Varianz

7.2 Ausnutzung der Diskontinuität der Störsignale

In vielen Spezialfällen haben die Störgeräusche in der Fertigung besondere Eigenschaften, die man ausnutzen kann, um die Vorverarbeitung zu vereinfachen. Beispielsweise hat der Schallpegel der zu untersuchenden Maschine über einen längeren Zeitraum einen konstanten Wert, während die wichtigsten Störquellen nur intermittierend auftreten. Dann ist es nicht nötig, das originale Motorsignal aus der gestörten Aufnahme zu rekonstruieren. Vielmehr reicht es aus, die Zeitpunkte zu finden, an denen keine Störung auftritt. Nur zu diesen Zeitpunkten ist die Klassifikation auszuführen.

Im folgenden wird ein Algorithmus vorgestellt, der die Quellenanzahl auf Basis der Eigenvektoren der (räumlichen) Kovarianzmatrix für jeden Zeitpunkt bestimmt. In der Systematik der Signalverarbeitung basiert er auf einem Subspace-Ansatz und wird darum im folgenden als Subspace-Verfahren bezeichnet. Ist nur eine Quelle vorhanden, können die Daten für die Klassifikation verwendet werden, da der Motor

das einzige kontinuierliche Geräusch ist. Ein ähnlicher Algorithmus wurde von van Rooijen *et al.* (2002) vorgeschlagen und ist für die Adaption eines Beamformers auf die Sprachsignalpausen bestimmt. Allerdings ist der von ihnen verwendete Algorithmus relativ komplex in Hinblick auf die verwendeten Parameter, so daß in dieser Arbeit ein einfacherer vorgestellt wird, der ähnliche Ergebnisse bringt.

7.2.1 Schätzung des Signalunterraumes

Wie schon in den vorangegangenen Kapiteln wird von dem Signalmodell ausgegangen, das in Kap. 4.1 eingeführt wurde. Zusätzlich wird für diesen Algorithmus angenommen, daß die Anzahl der Observationskanäle N die Anzahl der gerichteten Quellen M übersteigt. Dafür wird ein großer Anteil diffusen Schalls zugelassen. Es muß lediglich garantiert werden, daß seine Leistung nicht höher als die der gerichteten Quellen ist. Der Observationsvektor \mathbf{x}^b wird wieder mit einer STFT in den Frequenzbereich transformiert, so daß man die schmalbandigen Signale \mathbf{x} erhält. Damit gilt für jeden Frequenzabschnitt: $\mathbf{x} = \mathbf{H}^{N \times M} \mathbf{s}^{M \times 1} + \mathbf{n}$. Dabei seien s_1 wieder das Nutzsignal s_{dut} (in diesem Fall eine Waschmaschine) und s_2, \dots, s_M die intermittierenden Störquellen.

Die Kovarianzmatrix der (mittelwertfreien) Observationen ist

$$\begin{aligned} \mathbf{R}_x &= E\{\mathbf{x}\mathbf{x}^H\} = E\{(\mathbf{H}\mathbf{s} + \mathbf{n})(\mathbf{H}\mathbf{s} + \mathbf{n})^H\} \\ &= \mathbf{H}\mathbf{R}_s\mathbf{H}^H + E\{\mathbf{n}\mathbf{n}^H\} + \mathbf{H}E\{\mathbf{s}\mathbf{n}^H\} + E\{\mathbf{n}\mathbf{s}^H\}\mathbf{H}^H \\ &= \mathbf{H}\mathbf{R}_s\mathbf{H}^H + E\{\mathbf{n}\mathbf{n}^H\} = \mathbf{R}_{xs} + \mathbf{K}_n, \end{aligned}$$

wenn der nicht gerichtete Anteil und die Signale unkorreliert sind ($\forall i, j : E\{s_i \cdot n_j^*\} = 0$). Für die Bestimmung der Kovarianzmatrix des diffusen Schalls \mathbf{K}_n wird angenommen, daß dieser isotrop ist (vgl. Kap. 3.1.2). Dann existieren keine Kreuzkovarianzen zwischen den Observationskanälen und der Schalldruck ist überall gleich, wenn die gerichteten Quellen fehlen. Damit ist \mathbf{K}_n eine mit der Leistung des Diffusschalls σ_n gewichtete Einheitsmatrix.

$$\mathbf{K}_{n,\text{isotrop}} = \sigma_n^2 \mathbf{I} \quad (7.2)$$

Die Kovarianzmatrix der gerichteten Quellen \mathbf{R}_{xs} hat den Rang M , weil sie aus genau M unabhängigen Gleichungen zusammengesetzt ist, und damit M von Null verschiedene Eigenwerte $\lambda_{y,m}$ und hermitesche Eigenvektoren $\mathbf{u}_{y,m}$ besitzt. Diese spannen das Bild⁵ von \mathbf{R}_{xs} auf, in dem die Signalvektoren $\mathbf{h}_m \cdot s_m$ liegen. Nun definiert man $N - M$ weitere Vektoren \mathbf{v}_n , die aus dem Kern⁶ von \mathbf{R}_{xs} stammen und die untereinander hermitisch sind. Mit der Erweiterung der Diagonalmatrix der Eigenwerte zu $\mathbf{\Lambda}_y = \text{diag}[\lambda_{y,1}, \dots, \lambda_{y,M}, \mathbf{0}^{N-M \times 1}]$ und der Zusammenfassung der Vektoren $\mathbf{u}_{y,m}$ und \mathbf{v}_n zu den Matrizen \mathbf{U} und \mathbf{V} ergibt sich folgende Eigenwertzerlegung

$$\mathbf{R}_{xs} = \begin{bmatrix} \mathbf{U}^H \\ \mathbf{V}^H \end{bmatrix} \cdot \mathbf{\Lambda}_y \cdot [\mathbf{U} \quad \mathbf{V}].$$

⁵Bild(\mathbf{A}) = $\{\mathbf{y} \in \mathbb{R}^N : \mathbf{y} = \mathbf{A}\mathbf{x}$ für ein beliebiges $\mathbf{x} \in \mathbb{R}^N\}$, engl. range.

⁶Kern(\mathbf{A}) = $\{\mathbf{x} \in \mathbb{R}^N : \mathbf{0} = \mathbf{A}\mathbf{x}\}$, engl. nullspace.

Da $\mathbf{K}_{n,\text{isotrop}}$ eine Einheitsmatrix ist, hat sie N identische Eigenwerte σ_n^2 , und jeder beliebige Einheitsvektor ist ein Eigenvektor. Damit sind auch die Vektoren $\mathbf{u}_{y,m}$ und \mathbf{v}_n Eigenvektoren und es gilt:

$$\begin{aligned}\mathbf{R}_x &= \begin{bmatrix} \mathbf{U}^H \\ \mathbf{V}^H \end{bmatrix} \cdot \mathbf{\Lambda}_y \cdot [\mathbf{U} \ \mathbf{V}] + \begin{bmatrix} \mathbf{U}^H \\ \mathbf{V}^H \end{bmatrix} \cdot \sigma_n^2 \mathbf{I} \cdot [\mathbf{U} \ \mathbf{V}] \\ &= \begin{bmatrix} \mathbf{U}^H \\ \mathbf{V}^H \end{bmatrix} \cdot (\mathbf{\Lambda}_y + \sigma_n^2 \mathbf{I}) \cdot [\mathbf{U} \ \mathbf{V}].\end{aligned}\quad (7.3)$$

Das heißt, daß $[\mathbf{U} \ \mathbf{V}]$ die Matrix der Eigenvektoren von \mathbf{R}_x ist. \mathbf{U} und \mathbf{V} spannen zwei orthogonale Unterräume auf, wobei die Signale nur in dem Raum liegen können, der von Vektoren von \mathbf{U} aufgespannt wird. Seine Dimension entspricht der Anzahl der aktiven Quellen.

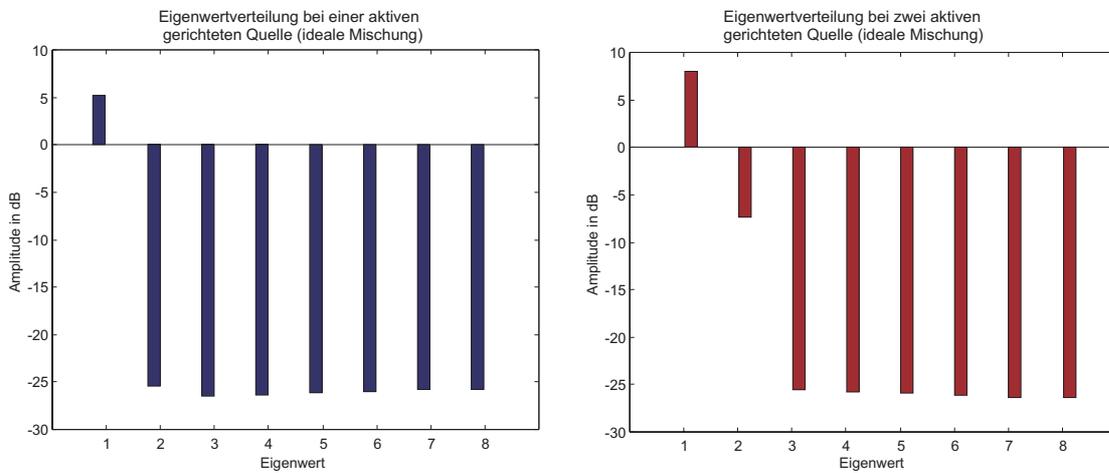


Abb. 7.8: Eigenwertverteilung der Kovarianzmatrix mit einer bzw. zwei aktiven Quellen in dB unter idealen Bedingungen (unkorrelierter diffuser Schall mit deutlich geringer Leistung als die gerichteten Quellen, Punktquellen)

Die von Null verschiedenen Eigenwerte von \mathbf{R}_{xs} liegen zwischen dem minimalen und maximalen Rayleigh-Koeffizienten von \mathbf{R}_{xs} . Da Eigenwerte von \mathbf{R}_x aus der Summe von $\lambda_{y,m}$ bestehen, sind die zu \mathbf{U} gehörigen Eigenwerte wenigstens den minimalen Rayleigh-Koeffizienten von \mathbf{R}_{xs} größer als die zu \mathbf{V} gehörigen. So kann man die beiden Räume nach der Größe der Eigenwerte leicht unterscheiden. Abb. 7.8 illustriert dieses Verhalten: In den Teilabbildungen (a) und (b) sind jeweils eine oder zwei Quellen aktiv, also sind ein bzw. zwei Eigenwerte größer als die Leistung des diffusen Schalls (σ_n). Demzufolge hilft ein einfaches Zählen, um die Anzahl der aktiven Quellen zu ermitteln (siehe auch Abb. 7.10a).

Ist der diffuse Schall nicht isotrop (z.B. weil er nicht stationär ist oder gewisse Vorzugsrichtungen vorhanden sind), ist seine Kovarianzmatrix keine gewichtete Einheitsmatrix mehr. Damit ist (7.3) nicht mehr anwendbar und die Eigenwertzerlegung allein führt auf keine klar getrennten Unterräume. In der Literatur über Subspace-Verfahren wird darum eine generalisierte Eigenwertzerlegung⁷ vorgeschlagen, die die

⁷In der deutschsprachigen mathematischen Literatur wird dieses Problem meist als „Eigenwert-

Kenntnis von \mathbf{K}_n erfordert. Das generalisierte Eigenwertproblem lautet

$$\mathbf{R}_x \mathbf{u} = \lambda \mathbf{K}_{\min} \mathbf{u}.$$

Darin ist \mathbf{K}_{\min} die normalisierte Kovarianzmatrix⁸ des Diffusschalls mit $\text{tr}(\mathbf{K}_{\min}) = N$, die man erhält, wenn man die Hauptleistung des Rauschens vorzieht: $\mathbf{K}_n = \sigma_n \mathbf{K}_{\min}$.

$$\mathbf{R}_x \mathbf{u} = \mathbf{R}_{xs} \mathbf{u} + \sigma_n \mathbf{K}_{\min} \mathbf{u} = \lambda \mathbf{K}_{\min} \mathbf{u}$$

Die $N - M$ kleinsten Eigenwerte λ_{\min} gehören zum Kern von \mathbf{R}_{xs} und haben den Wert σ_n .

$$\mathbf{R}_{xs} \mathbf{u} + \sigma_n \mathbf{K}_{\min} \mathbf{u} = \lambda \mathbf{K}_{\min} \mathbf{u} \quad \rightarrow \quad \sigma_n \mathbf{K}_{\min} \mathbf{u} = \lambda \mathbf{K}_{\min} \mathbf{u} \quad \rightarrow \quad \sigma_n = \lambda_n$$

Die anderen Eigenwerte sind, wie man leicht erkennt, die Summe aus σ_n und dem Rayleigh-Quotienten von \mathbf{R}_{xs} und \mathbf{u} . Dieser liegt zwischen dem kleinsten und größten Eigenwert von \mathbf{R}_{xs} und ist immer positiv, so daß diese generalisierten Eigenwerte größer als die erstgenannten sind, so daß man anhand der Größe der generalisierten Eigenwerte die Dimension des Signal-Unterraums und damit die Anzahl der Quellen leicht bestimmen kann.

In der Praxis ist das jedoch schwierig, weil dazu eine genaue Kenntnisse der Kovarianzmatrix des diffusen Schalls benötigt wird. Die hohe Anforderung an die Genauigkeit entsteht dadurch, daß das generalisierte Eigenwertproblem $\mathbf{R}_{xs} \mathbf{u} = \lambda \mathbf{K}_n \mathbf{u}$ auf das einfache Eigenwertproblem $\mathbf{K}_n^{-1} \mathbf{R}_{xs} \mathbf{u} = \lambda \mathbf{u}$ zurückgeführt wird und die Kovarianzmatrix nahezu eine Einheitsmatrix ist, so daß sich kleine Fehler bei der Invertierung somit erheblich auswirken. Da hier nur eine ungleichzeitige Schätzung von \mathbf{K}_n möglich ist, ist eine solche Genauigkeit nicht erreichbar, weil nicht von einer völligen Stationarität ausgegangen werden kann.

7.2.2 Bestimmung der Dimensionalität des Signalunterraums

Die Basis des Verfahren ist, die gesamte Eigenwertverteilung von \mathbf{R}_x zu berücksichtigen, denn die Auszählung der dominanten Eigenwerte ist allein nicht aussagekräftig. Die Gründe sind die folgenden:

- Die Annahme, daß \mathbf{K}_n eine gewichtete Einheitsmatrix ist, trifft nicht immer zu. Damit gilt (7.3) nicht. Leider ist die Lösung des generalisierten Eigenwertproblems aus den genannten Gründen ebenfalls nicht möglich.
- Der Anteil der gerichteten Quellen an den Observationen ist stark korreliert, so daß die Eigenwerte von \mathbf{R}_{xs} sehr stark gespreizt sind. Auch der diffuse Schall ist korreliert, so daß sich seine Leistung ebenfalls nicht gleichmäßig

problem" bezeichnet, während das hier als „Eigenwertproblem" bezeichnete Problem $\mathbf{R}\mathbf{u} = \lambda\mathbf{u}$ „spezielles Eigenwertproblem" genannt wird, vgl. Zurmühl und Falk (1992). Die hier verwendete Nomenklatur entspricht der meist englischsprachigen Signalverarbeitungsliteratur, z.B. Therrien (1992).

⁸Ohne weitere Untersuchung werden \mathbf{K}_n und damit \mathbf{K}_{\min} als invertierbar angenommen.

auf alle Eigenwerte verteilt. Als Konsequenz unterscheiden sich die kleinsten Eigenwerte des Signalunterraumes nicht mehr deutlich von denen des diffusen Schalls, wie in Abb. 7.9 sichtbar.

- Das Signalmodell in Kap. 4.1 geht von Punktquellen aus. In vielen Fällen, z.B. bei Waschmaschinen, wird diese Annahme verletzt.

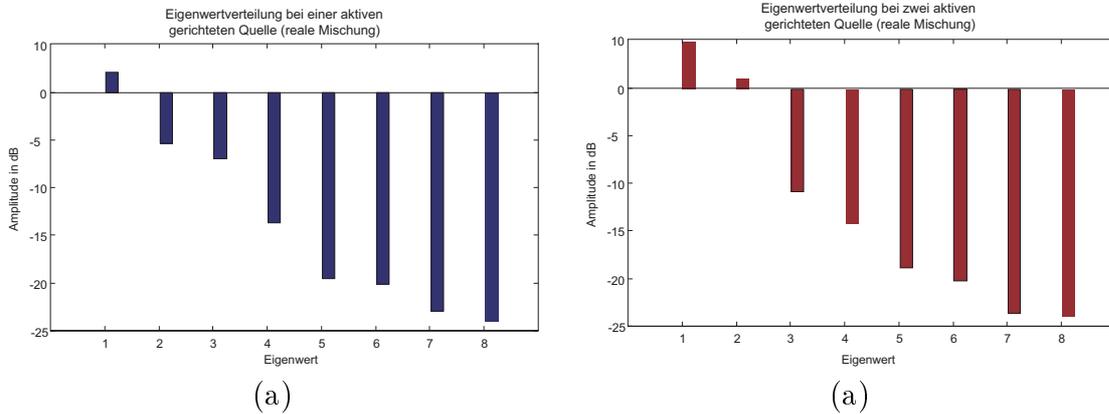


Abb. 7.9: Eigenwertverteilung der Kovarianzmatrix mit einer (a) bzw. zwei (b) aktiven Quellen in dB für eine reale Messung unter den in Kap. 7.2.3 beschriebenen Bedingungen

Obwohl sich die Anzahl der Quellen nicht mehr ohne weiteres ablesen lässt, bilden die Verteilungen der Eigenwerte charakteristische Cluster in Abhängigkeit von der Anzahl der Quellen. Diese kann man mit einem statistischen oder anderen Klassifikator (Fukunaga (1990)) voneinander unterscheiden.

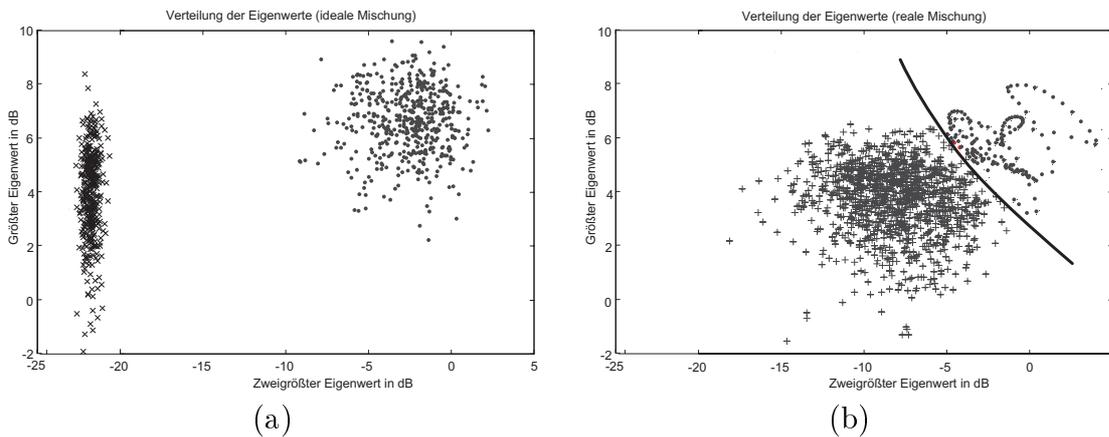


Abb. 7.10: Streudiagramm der beiden größten Eigenwerte für eine (x) oder zwei (o) aktive Quellen; (a) ideale Simulation und (b) reale Messung

Von van Rooijen *et al.* (2002) wurde eine „Support Vector Machine“ (vgl. Haykin (1994)) vorgeschlagen. Sie stammt aus der Theorie der neuronalen Netze und ist nach ihrer Struktur ein nichtlinearer Klassifikator. Knaak und Filbert (2003) verwenden einen Mahalanobis-Klassifikator, einen sehr einfachen statistischen Klassifikator,

der ebenfalls zufriedenstellende Ergebnisse liefert, wie im folgenden gezeigt wird. Er hat den Vorteil, daß er durch seine Linearität deutlich weniger komplex ist und damit weniger Parameter braucht, die eingestellt werden müssen. Der Mahalanobis-Klassifikator hat nur die Mittelpunkte und die Kovarianzmatrizen der Klassen als Parameter, während die Support Vector Machine den gesamten Bereich des Supportvektors durch Stützstellen abbilden muß.

Für die Auswahl des Klassifikators wurde versucht, den einfachsten zu finden, der das Problem löst. Der noch einfachere Mittelpunkts-Abstands-Klassifikator mußte verworfen werden, weil er implizit die Annahme gleicher Kovarianzmatrizen besitzt, welche hier jedoch sehr stark variieren, wie in Abb. 7.10b erkennbar ist. Darum wurde der Mahalanobis-Klassifikator gewählt, der diese Eigenschaft berücksichtigt.

Als Eingangsvektor für den Klassifikator wird der Vektor der Eigenwerte von \mathbf{R}_x genutzt: $\mathbf{f} = [\lambda_{x,1}, \dots, \lambda_{x,1}]$. Die beiden Klassen „nur ein Signal vorhanden“ (Klasse 1) und „zwei Signale vorhanden“ (Klasse 2) formen zwei Cluster mit den Mittelpunkten \mathbf{m}_1 und \mathbf{m}_2 (vgl. Abb. 7.10b). Die Abbildung zeigt, daß auch bei realen Signalen der zweite Eigenwert von der Anzahl der Signale abhängig ist: Ist nur eine Quelle aktiv, ist er im Mittel deutlich kleiner. Allerdings ist die Streuung so groß, daß eine Klassifikation nur auf Basis des zweiten Eigenwertes nicht erfolgreich sein wird.

Die Ausdehnung der beiden Cluster wird durch die Intraklassenkovarianzen \mathbf{C}_i beschrieben. Der Mahalanobis-Abstand ist der (euklidische) Abstand eines Vektors zum Mittelpunkt der Klasse bezogen auf die Streuung der Klasse.

$$\varrho^2(\mathbf{f}, i) = (\mathbf{f} - \mathbf{m}_i)^T \mathbf{C}_i^{-1} (\mathbf{f} - \mathbf{m}_i) \quad (7.4)$$

Der Vektor \mathbf{f} gehört der Klasse i an, wenn gilt

$$\varrho^2(\mathbf{f}, i) < \varrho^2(\mathbf{f}, j) \quad \forall j \neq i \in I. \quad (7.5)$$

7.2.3 Untersuchungsbedingungen und Algorithmus

Zum Test dieses Verfahrens wurde ein Mikrofonarray mit acht Mikrofonen entsprechend Abb. 7.11 neben der Linie einer Waschmaschinenproduktion aufgebaut. Die Abstände der Mikrofone variierten zwischen 4 cm und 10 cm. Die Mikrofone waren abgegliche Elektretmikrofone (vgl. Heckl und Müller (1994)) und wiesen einen Verstärkungsunterschied von 1,5 dB auf, was für diesen Algorithmus ausreichend ist. Die Abtastrate betrug 44 kHz, durch Sample & Hold-Bausteine ist eine synchrone Abtastung aller Kanäle garantiert.

Von den Waschmaschinen wurde die Rückwand, wie in Abb. 7.11 dargestellt, entfernt, um die Motoren- und Störgeräusche besser aufzunehmen. Zu den Fehlern, die akustisch wirksam sind, gehören schlecht montierte Teile, wie z.B. ein loser Reibdämpfer. Diese erzeugen ein klickendes Geräusch während des Schleuderns mit einer Unwucht, weil sie gegen die nun stark vibrierende Trommel schlagen. In unmittelbarer Nähe zum Meßort an der Güteprüfung befinden sich Hubtische, mit deren Hilfe die Waschmaschinen für die Funktions- und die manuelle akustische Güteprüfung ausgesondert werden. Dieser gibt beim Anschlag ein stark impulshaltiges Geräusch

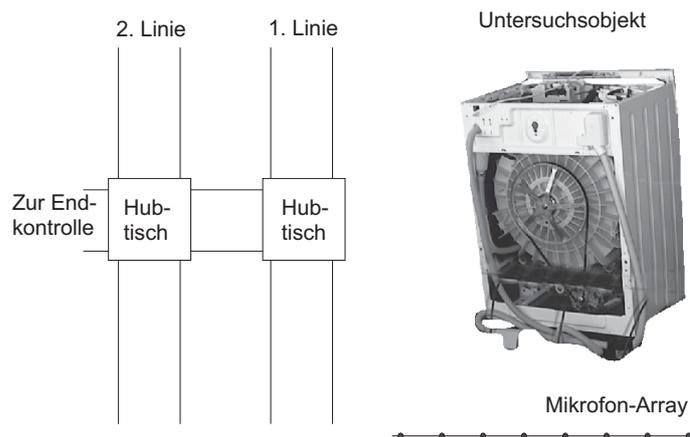


Abb. 7.11: Messaufbau für die Untersuchung des Subspace-Algorithmus in der Fertigungshalle für Waschmaschinen.

von sich. Im Hintergrund waren weitere Maschinen, Gebläse und gelegentlich Rufe hörbar.

Mit diesem Aufbau wurden Aufnahmen von 15 s Dauer von mehreren Maschinen mit und ohne Fehler gemacht. In dieser Zeit bewegte sich der Hubtisch bei fast allen Aufnahmen mehrfach. Bei fünf Aufnahmen wurden die Zeitabschnitte, in denen der Hubtisch aktiv war, manuell mit „jammer“, die restlichen mit „quiet“⁹ bewertet. Nach der Transformation in den Frequenzbereich wird für jeden Zeitpunkt eine Kovarianzmatrix (einer aussagekräftigen Frequenzlinie \mathbf{f}^f) durch gleitende Mittelung über vorhergehende Frames geschätzt: $\mathbf{R}_x^f = \sum_n \mathbf{x}^f(t-n)\mathbf{x}^{fH}(t-n)$.

Die sechs größten Eigenwerte von \mathbf{R}_x^f bilden einen Merkmalsvektor¹⁰. Alle Merkmalsvektoren, deren Zeitabschnitte mit „jammer“ bewertet wurde, gehören zum Trainingsdatensatz für die Klasse „jammer“, alle anderen zum Datensatz der Klasse „quiet“. Damit standen 75 Merkmalsvektoren für das Training des Klassifikators zur Verfügung. Mit ihnen wurden Kovarianzmatrizen der Klassen und die Klassenmittelpunkte geschätzt. Die damit erzielten Ergebnisse werden in Kap. 8.4 beschrieben.

⁹Diese Bezeichnung lautet beabsichtigt nicht „silent“, weil erhebliche diffuse Schallanteile vorhanden waren, wie man an der Verteilung der Eigenwerte in Abb. 7.9 erkennen kann. Da sie jedoch nicht impulshaltig sind, stören sie die Klassifikation nicht und werden darum als irrelevant angesehen.

¹⁰Theoretisch eines beliebigen Frequenzbänder f . Hier wurde als aussagekräftiges Frequenzband $f=1$ kHz gewählt, weil alle Komponenten starke Leistungsanteile in diesem Frequenzbereich haben. Zur Erhöhung der Verlässlichkeit ist es ebenso, mehrere Frequenzbänder auszuwählen und daraus einen Merkmalsvektor zu formen.

Kapitel 8

Untersuchung der neuen Verfahren und Ergebnisse

Der Schwerpunkt dieses Kapitels liegt auf der Bewertung der vorgestellten und neu eingeführten Algorithmen in Hinblick auf die technische Diagnose (Kap. 8.3 und Kap. 8.4). Das wirft zunächst die Frage nach geeigneten Bewertungsmaßen auf, mit denen der Nutzen der Rekonstruktionsverfahren angegeben werden kann (Kap. 8.2). Die Ziele der Rekonstruktion bestehen darin, die Neben- oder Störgeräusche zu unterdrücken und gleichzeitig das Nutzsignal nicht zu beeinflussen. Da die beiden Teilziele konträr sein können, muß man für jedes ein eigenes Bewertungsmaß definieren,

Die Unterdrückung der Nebengeräusche wird über die Veränderung des SIR (signal-to-interference-ratio, Signal-Stör-Verhältnis)¹ erfaßt. Es ist das Verhältnis zwischen der Energie des im Signal enthaltenen Nutzanteils zur Energie des Störanteils. Problematisch ist in vielen Fällen die Schätzung dieser Anteile, so daß in Abhängigkeit von der Art der Schätzung unterschiedliche Definitionen des SIR existieren. Die hier verwendete Definition wird in Kap. 8.2.1 genauer besprochen.

Leider ist es unvermeidlich, daß bei der Unterdrückung von Nebengeräuschen das Nutzsignal unvorteilhaft beeinflusst wird. Das ist natürlich unerwünscht, so daß die Maximierung des SIR allein nicht immer zum optimalen Ergebnis führt. Die Qualität des Nutzsignals kann nur in Hinblick auf die Anwendung bewertet werden, weil bestimmte Veränderungen des Signals in der einen Anwendung desaströs sein können, während die gleiche Änderung in der anderen Anwendung vernachlässigbar ist. In der Sprachverarbeitung, dem größten Anwendungsgebiet für die Rekonstruktion akustischer Signale, ist sie äquivalent mit der Sprachverständlichkeit, welche je nach Einsatzgebiet mit formellen und informellen Hörtests oder der Erkennungsrate eines

¹In der Literatur sind ebenfalls die Größen SNR (signal-to-noise-ratio) oder SINR (signal-to-interference-&-noise-ratio) gebräuchlich. Die Unterschiede resultieren aus den unterschiedlichen Definitionen der Störung. Einige Autoren (z.B. Drews (1999)) bezeichnen jegliche Störung als Rauschen, so daß sie den SNR als Bewertungsmaß verwenden. Häufiger ist Differenzierung zwischen Störquellen („interference“ i.e.S.) und dem Rauschen des Meßsystems („noise“), die im Namen SINR ihren Niederschlag findet. In dieser Arbeit wird der SIR verwendet, weil das Rauschen der Mikrofone unter 20 dB liegt und im sehr tiefen Frequenzbereich angesiedelt ist, so daß es vernachlässigbar ist.

automatischen Spracherkenners bestimmt wird. Während erstere in der technischen Diagnose unsinnig sind, ist das Gegenstück zur Erkennungsrate die Klassifikationsrate, da das ultimative Ziel der technischen Diagnose in einer möglichst präzisen Klassifikation des Maschinenzustands anhand der akustischen Signale besteht.

Diese Bewertung hat jedoch einige Nachteile. Ein Nachteil besteht darin, daß die Auswahl des Klassifikators und besonders der Merkmale subjektiv und problem-spezifisch ist (Parsons (1987)). Dadurch verliert man die Allgemeingültigkeit der Aussage. Weiterhin können die Merkmale unterschiedlich robust gegenüber Störungen sein, so daß ähnliche Störungen ein unterschiedliches Resultat ergeben können. Wie in Kap. 1 gezeigt, können Klassifikation und Signalschätzung als Parameterschätzungen angesehen werden, so daß ein guter (robuster) Klassifikator trotz nicht unterdrückter Nebengeräusche gute Resultate erzielen kann, während ein sehr sensibler Klassifikator bereits bei kleinen Verzerrungen unzureichend arbeitet.

Diese Einschränkungen anerkennend, wird dennoch eine Bewertung anhand der Klassifikation vorgenommen. Jedoch soll der Einfluß der Klassifikation und Merkmalsauswahl auf das Ergebnis möglichst gering gehalten werden. Dazu werden in Kap. 8.2.2 zunächst die Merkmale aus einer großen Merkmalsbibliothek ausgewählt, mit denen sich die ungestörten Signale am besten trennen lassen. Um den Einfluß des Klassifikators gering zu halten, wird die Trennbarkeit dieser Merkmale über das Spurkriterium (Ruske (1988)) und einen wenig problemangepaßten Klassifikator beschrieben. Ein Vergleich der Trennbarkeit dieser Merkmale für gestörte und rekonstruierte Signale gibt Aufschluß über die Qualitätsverbesserung durch die Rekonstruktion²

Die Algorithmen wurden in mehreren Testumgebungen mit definierten Bedingungen getestet. Als Testumgebungen werden sowohl Räume, in denen Aufnahmen gemacht wurden, als auch Datensätze mit den Raumimpulsantworten realer Räume angesehen, mit denen „reale“ Mischungen³ künstlich erzeugt werden können. Die ICA-basierten Algorithmen (inkl. der adaptiven Beamformer) wurden mit den drei Testumgebungen ausgiebig getestet, die in Kap. 8.1 beschrieben werden. Bei zwei Testumgebungen handelte es sich um Datensätze mit Impulsantworten; die dritte war ein realer Raum. Mit diesen Tests wurde die sehr gute Performance der ICA-Algorithmen sichtbar. In allen Fällen konnte eine signifikante Verbesserung des SIR und der Klassifizierbarkeit der Signale - auch in Räumen mit Nachhall - nachgewiesen werden.

Die Bewertung des Subspace-Verfahrens erfolgte hingegen mit Aufnahmen in einer Fabrikhalle. Der experimentelle Aufbau wurde schon in Kap. 7.2.3 beschrieben. Damit konnte dieser Algorithmus sogar in seiner Anwendungsumgebung erfolgreich

²In der Spracherkennung wird diese Vorgehensweise in der Spracherkennung „Orakel“ genannt: Die eigenen Ergebnisse werden mit einem als optimal angesehenen, jedoch nicht notwendigerweise perfekten System verglichen

³Diese Mischungen werden in der Literatur praktisch immer als reale Mischungen bezeichnet, obwohl sie rechnerisch erfolgten, weil sie reale Impulsantworten verwenden. Da die Meßsysteme hinreichend rauscharm waren, hängt es nur von der Linearität des Meßsystems ab, ob diese Bezeichnung gerechtfertigt ist. Ist es linear, ist es in der Tat unerheblich, ob die gefilterten Signale vor der Messung akustisch oder danach rechnerisch gemischt werden.

getestet werden, weshalb der Test in einer definierten Umgebung entfiel.

8.1 Beschreibung der Testumgebungen

Die Algorithmen wurden zum einen mit gestörten Motorsignalen untersucht, die in einem speziell vermessenen Raum⁴ mit variabel einstellbarer Nachhallzeit gemessen wurden. Zum anderen wurden gestörte Signale verwendet, die über eine Datenbank mit (realen) Raumimpulsantworten generiert wurden. Die Abmessungen des Meßraumes sind in Abb. 8.1 dargestellt.

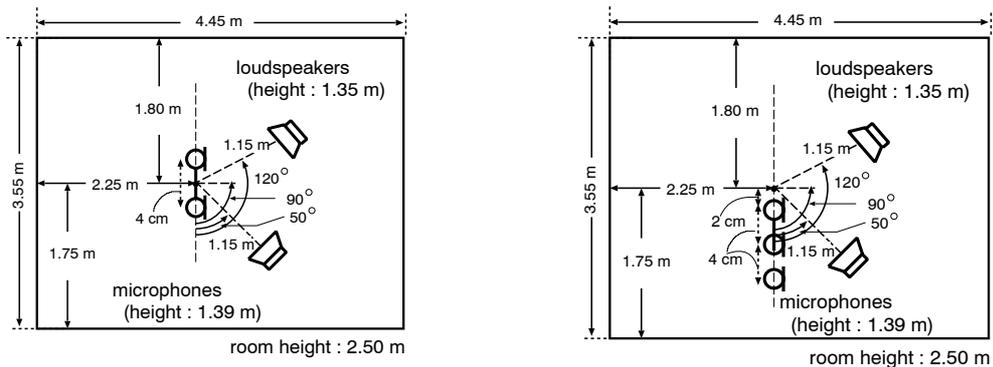


Abb. 8.1: Abmessungen des Meßraumes und Anordnung des Mikrofon-Arrays mit zwei (a) und drei (b) Mikrofonen bzw. Quellen

In diesem Raum befindet sich ein lineares Mikrofonarray mit insgesamt acht Mikrofonen in unterschiedlichen Abmessungen. In der Mitte haben die Mikrofone einen Abstand von 1 cm, während die äußeren einen Abstand von 4 cm besitzen. Damit können verschiedene Mikrofonabstände und somit die Frequenzabhängigkeit des Arrays untersucht werden. Die Mikrofone sind aufeinander abgeglichen, der Verstärkungsunterschied beträgt weniger als 0.5 dB, der Signal-Rausch-Abstand ca. 20-21 dB.

Als Signalquellen dienten Lautsprecher, deren Position variierbar war und mit denen die akustischen Signale abgespielt wurden. Als Nutzsignale dienten die Aufnahmen aus Kap. 7.1 von den dort beschriebenen Kleinmotoren mit unterschiedlichen Fehlerzuständen. Die Störsignale waren Aufnahmen aus einer Werkhalle und sind ebenfalls in Kap. 7.1 beschrieben. Somit stand eine sehr gute Meßumgebung zur Verfügung.

In den Experimenten wurden zwei bzw. drei Mikrofone benutzt. Für die Algorithmen, die auf die Auswertung von zwei Observationssignalen begrenzt sind (siehe 8.2), wurden nur zwei Mikrofone im Abstand von 4 cm verwendet, und das Motorengeräusch wurde von einem Lautsprecher, der Industriegeräusche abspielte, gestört. Abb. 8.1a zeigt diesen Aufbau. Ermöglichen die Algorithmen die Auswertung von

⁴der bei den Communication Science Labs der NTT Corp. in Kyoto zur Verfügung stand

drei Observationen, wird das Mikrofonarray um ein Mikrofon erweitert und das Motorsignal durch zwei Lautsprecher gestört (Abb. 8.1b).

Weiterhin wurde eine Datenbank mit Impulsantworten eines nahezu hallfreien Raumes verwendet. Diese hat zwar nur theoretischen Wert, kann aber zu einer besseren Einordnung der Algorithmen beitragen. Alle verwendeten Testumgebungen sind mit ihrem Kürzel und den Nachhallzeiten im Tab. 8.1 zusammengefaßt.

Kürzel	Art der Messung	Nachhallzeit
DASBox	Meßraum nach Abb. 8.1	150 ms
NST	voraufgenommene Impulsantworten	ca. 15 ms

Tabelle 8.1: Testumgebungen mit den Nachhallzeiten

Untersuchungen in einer Fertigungshalle für Kleinmotoren wurden nicht vorgenommen. Voruntersuchungen ergaben, daß der Schallpegel der Umgebungsgeräusche in der zur Verfügung stehenden Halle bei etwa 60 dB liegt, während die Motoren einen Schallpegel von ca. 40 dB, haben. Bei einem Pegelunterschied von ca. 20 dB kann das menschliche Ohr das leisere Geräusch kaum wahrnehmen und erst recht keine Entscheidungen darauf basierend fällen. Da die menschliche Leistung oft als Obergrenze der technischen Möglichkeiten angesehen wird, ist eine signifikante Verbesserung der Klassifikationsrate nicht zu erwarten.

Da leider keine Produktionsstätten von Kleinmotoren zur Verfügung standen, in denen die Motoren und die Nebengeräusche etwa gleich laut gewesen wären, können die im folgenden vorgestellten Ergebnisse nur extrapoliert werden. Sie bestätigen aber dennoch das erhebliche Potential der mehrkanaligen Signalverarbeitung.

Für den Spezialfall, daß die Nebengeräusche intermittierend auftreten, konnte der in Kap. 7.2 vorgestellte Subspace-Algorithmus erfolgreich in einer Fertigungshalle eingesetzt werden. Der experimentelle Aufbau wurde bereits dort (Kap. 7.2.3) beschrieben, weil er für die Ableitung des konkreten Algorithmus nötig war.

8.2 Bewertungsmaße für die Rekonstruktion

Das primäre Maß für die Bewertung eines Algorithmus zur Störunterdrückung ist der SIR (Signal-to-Interference-Ratio). Er beschreibt die Energie des Nutzsignals im Verhältnis zu der des im Signal verbliebenen Störanteils. Dieses Maß ist energiebasiert und sagt damit nichts über die unvermeidliche Veränderung der Signale durch die Verarbeitung aus. Eine extreme Unterdrückung der Nebengeräusche kann eine derartig heftige Verzerrung zur Folge haben, daß die Sprachverständlichkeit im Falle der Sprachsignalverarbeitung und die Ausprägung charakteristischer Merkmale für die Fehlerdiagnose sinkt. Darum wird meistens ein weiteres Maß angegeben, mit dem Qualität des Nutzsignals bestimmt wird. In dieser Arbeit ist es die Klassifizierbarkeit.

8.2.1 Maße für die Unterdrückung der Nebengeräusche (SIR)

Die Unterdrückung des Nebengeräusches wird über den SIR in (8.1) bestimmt, darin ist y_{dut} der Anteil des Maschinengeräuschs und des y_s der Anteil der Nebengeräusche am Ausgangssignal.

$$\text{SIR} = 10 \log_{10} \frac{\text{E}\{y_{\text{dut}}^{b^2}\}}{\text{E}\{y_s^{b^2}\}} \quad (8.1)$$

Das Problem an dieser Definition ist die Bestimmung der Signalanteile im Ausgangssignal. Ein weit verbreitetes Maß ist der Segment-SIR von Noll (1974). Er bestimmt die Erwartungswerte, indem er aus den Ein- und Ausgangssignalen eine Anzahl von L_a Segmenten der Länge K_a bildet und damit den SIR über folgende Gleichung

$$\text{SIR}_{\text{seg}} = 10 \log_{10} \frac{1}{L_a} \sum_{l=0}^{L_a} \frac{\sum_{lK_a+1}^{(l+1)K_a} s_{\text{dut}}^{b^2}(k)}{\sum_{lK_a+1}^{(l+1)K_a} (y_1^{b^2}(k) - s_{\text{dut}}^{b^2}(k))^2}$$

berechnet. Der Gebrauch des Segment-SIR ist zur Beurteilung von Geräuschreduktionsverfahren allerdings problematisch: Zeitverschiebungen und Amplitudenunterschiede zwischen dem von der Maschine abgestrahlten Signal und dem störreduzierten Ausgangssignal verfälschen den Segment-SNR ausgesprochen stark, obwohl sie sich bei der auditiven Bewertung und einer Klassifikation kaum bemerkbar machen. Darum wurden weitere spektrale und parametrische Distanzmaße vorgeschlagen, die auf einem linearen Modell der Nutzsignalerzeugung beruhen und mit den Methoden der linearen Prädiktion berechnet werden, z.B. von Ahmed (1989). Diese Methoden sind an die Sprachsignalverarbeitung angelehnt und damit hier von geringem Nutzen.

Deshalb wird der SIR alternativ bestimmt, wobei die Tatsache ausgenutzt wird, daß in der Testumgebung die Stör- und die Nutzsignalquelle ausgeschaltet werden können. Zur Herleitung dieser SIR-Definition wird von zwei Quellen und Mikrofonen ausgegangen, obwohl sie allgemeingültig ist. Das erste rekonstruierte Signal ergibt sich gemäß (8.2). Abb. 8.2 zeigt den dazu gehörigen Signalfluß.

$$\begin{aligned} y_1^b &= w_{11}^b * x_1^b + w_{21}^b * x_2^b \\ x_1^b &= h_{11}^b * s_1^b + h_{21}^b * s_2^b \\ y_2^b &= h_{12}^b * s_1^b + h_{22}^b * s_2^b \\ \Rightarrow y_1^b &= \underbrace{(w_{11}^b * h_{11}^b + w_{21}^b * h_{12}^b)}_{y_{\text{dut}}} * s_1^b + \underbrace{(w_{11}^b * h_{21}^b + w_{21}^b * h_{22}^b)}_{y_s} * s_2^b \end{aligned} \quad (8.2)$$

Spaltet man (8.2) entsprechend den unteren Klammern auf, kann man die Einflüsse des Nutz- und des Störsignals leicht erkennen. Sie ergeben sich aus den in Abb. 8.3a/b dargestellten Signalflüssen. Diese Signalflüsse erhält man, indem man ein trainiertes System auf nur eine Quelle anwendet, d.h. die andere Quelle stumm stellt. Mit wechselseitigem Ausschalten kann man somit die Nutz- und Störanteile, also die Energien von y_{dut} und y_s , sehr leicht bestimmen und (8.1) direkt implementieren.

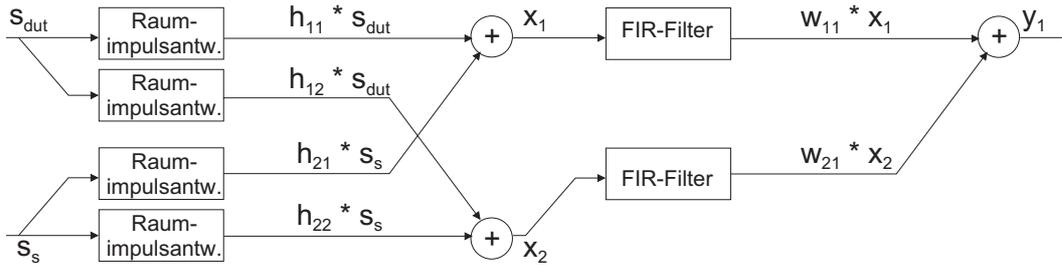


Abb. 8.2: Beitrag des Nutz- und des Störsignals zum rekonstruierten Signal

Durch eine Transformation der Signalanteile in den Frequenzbereich und die Bildung des Quotienten für jede Frequenz kann ein SIR in Abhängigkeit von der Frequenz definiert werden.

$$\text{SIR}_{\text{ref}}(f) = 10 \log_{10} \frac{E\{y_{\text{dut}}^2(f)\}}{E\{y_s^2(f)\}} \quad (8.3)$$

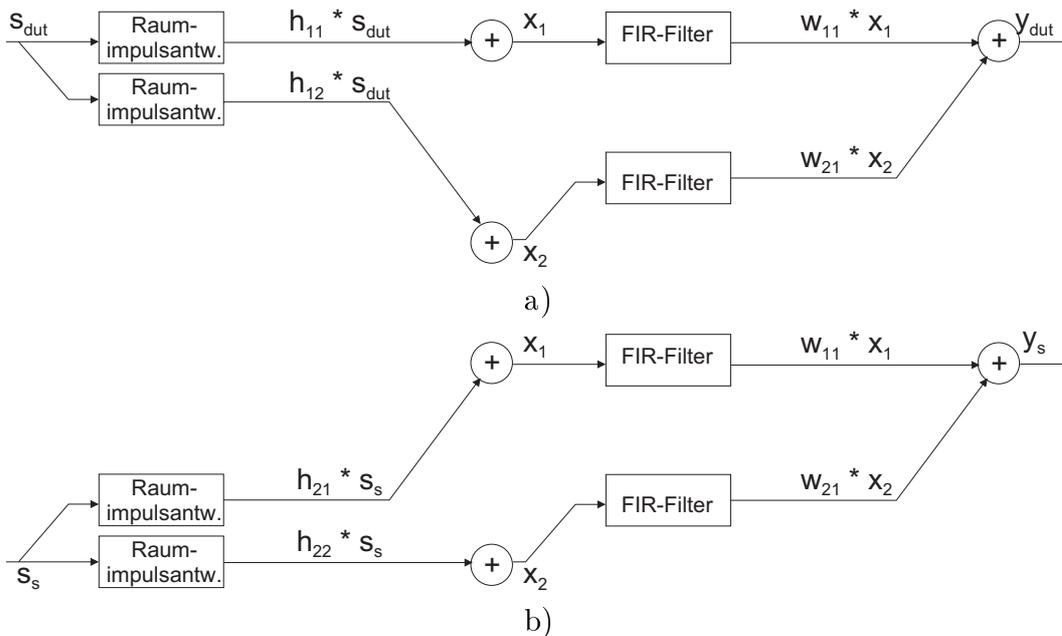


Abb. 8.3: Bestimmung der Anteile des Stör- (b) und Nutzsignals (a) am rekonstruierten Signal

Sind die Raumimpulsantworten durch Messungen bekannt, hat man zusätzlich die Möglichkeit, den SIR direkt aus den Impulsantworten zu bestimmen:

$$\text{SIR}_{\text{imp}} = 20 \log_{10} \frac{w_{11}^b * h_{11}^b + w_{21}^b * h_{12}^b}{w_{11}^b * h_{21}^b + w_{21}^b * h_{22}^b}.$$

Für jeden Frequenzabschnitt ergibt sich

$$\text{SIR}_{\text{imp}}(f) = 20 \log_{10} \frac{w_{11} h_{11} + w_{21} h_{12}}{w_{11} h_{21} + w_{21} h_{22}} = 20 \log_{10} \frac{\mathbf{w}_1^H \mathbf{h}_1}{\mathbf{w}_1^H \mathbf{h}_2},$$

was mit den Betrachtungen in Kap. 4 und 6 korrespondiert.

8.2.2 Maße für die Qualität des rekonstruierten Signals (Klassifizierbarkeit)

Um die Qualitätseinbußen des Nutzsignals zu beschreiben, wird die Trennbarkeit der rekonstruierten Signale im Verhältnis der Trennbarkeit der Originalsignale untersucht. Dazu wird ein Merkmalsatz extrahiert, der für die Originalsignale eine hohe Trennbarkeit sichert (Referenzklassifikation). Dann wird mit diesem Merkmalsatz die Trennbarkeit der gestörten bzw. rekonstruierten Signale getestet. Aus dem Unterschied (meist Verschlechterung) zur Referenzklassifikation kann man die Qualität des entsprechenden Signals und damit die Güte des Rekonstruktionsverfahrens ablesen. (Abb. 8.4)

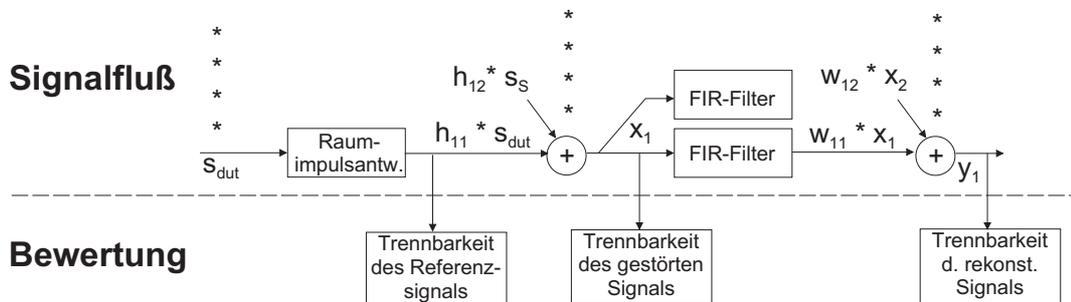


Abb. 8.4: Klassifikation an ausgewählten Punkten im Signalfluß von Abb. 8.2

Da die ICA-Verfahren mehrdeutig bezüglich einer Faltung sind, kann das Originalsignal am Lautsprecher nicht rekonstruiert werden. Um eine definierte Basis zu haben, versuchen die meisten Verfahren das Motorsignal am Ort des Mikrofons zu rekonstruieren, das dem Motor (oder in der Testumgebung dem Lautsprecher, der das Motorsignal abstrahlt) am nächsten ist. Darum werden auch als Referenzsignale für die optimale Klassifikation Motorsignale genutzt, die bereits mit der Raumimpulsantwort gefiltert sind (vgl. Abb. 8.4 und Abb. 8.5). Man erhält sie, indem man das Motorsignal abspielt und ohne zusätzliche Störung wieder aufnimmt.

Nach Ruske (1988) ist das ideale Gütemaß dann gegeben, wenn unmittelbar die zu erwartende Fehlerrate bei der Klassifikation angezeigt wird. Da die Trennbarkeit dann aber unmittelbar vom verwendeten Klassifikator abhängt, müßte im Prinzip das ganze Klassifikationssystem mit einbezogen werden. Da der eigentlich Klassifikator aber gerade nicht bewertet werden soll, werden zur Abschätzung darum oft der Mahalanobis-Abstand und das Spurkriterium verwendet⁵.

⁵Am besten wäre es, die Fehlerrate des optimalen Klassifikators (Bayes-Klassifikator) zu benutzen, was jedoch die genaue Kenntnis der Wahrscheinlichkeitsdichten der Klassen erfordert. Da diese nur bei sehr großen Stichproben hinreichend genau geschätzt werden können, müssen die obigen Abstandsmaße verwendet werden, die über zusätzliche Annahmen (normalverteilte Klassen und gleiche Auftrittswahrscheinlichkeit beider Klassen) aus dem optimalen Klassifikator hergeleitet werden können (siehe Ruske (1988)).

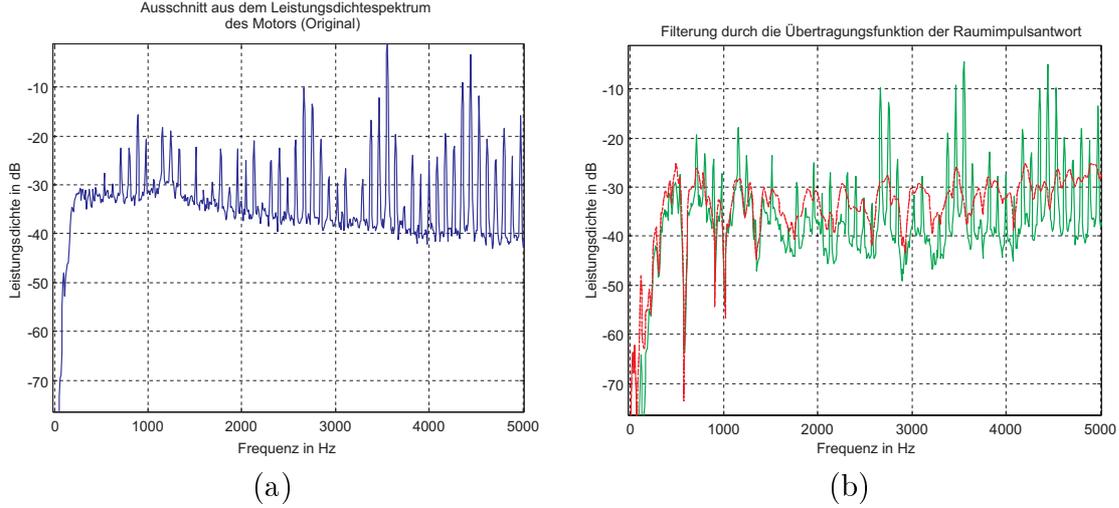


Abb. 8.5: Originalspektrum des Motors (a) und die Verfälschung durch die Filterung mit der Raumimpulsantwort (b): Durchgezogene Linie Signal am Ort des Mikrofons, gestrichelte Linie Übertragungsfunktion der Raumimpulsantwort

Für die Definition des Spurkriteriums wird davon ausgegangen, daß nach der Merkmalsextraktion jedem Motor k ein n -dimensionaler Merkmalsvektor \mathbf{f} zugeordnet ist.

$$s(k, t) \rightarrow \mathbf{f}_k \in \mathbb{R}^N$$

Der Fehlerzustand ξ_i erzeugt ein charakteristisches Schallsignal, so daß die Vektoren, die zu diesen Zuständen gehören, bei einer geeigneten Merkmalsauswahl einen kompakten Cluster im Merkmalsraum bilden. Der Cluster wird in erster Näherung als normalverteilt angesehen (mit \mathbf{m}_{ξ_i} , dem Vektor des Klassenmittelpunktes, und \mathbf{C}_{ξ_i} , der Kovarianzmatrix).

$$p(\mathbf{f}|\xi_i) = \frac{1}{(2\pi)^{N/2} |\mathbf{C}_{\xi_i}|^{1/2}} e^{-\frac{1}{2}(\mathbf{f}-\mathbf{m}_{\xi_i})^T \mathbf{C}_{\xi_i}^{-1} (\mathbf{f}-\mathbf{m}_{\xi_i})} \quad (8.4)$$

Das Spurkriterium ist der Abstand zweier Klassen im Merkmalsraum im Verhältnis zur Ausdehnung der Klassen. Den Abstand der Klassen voneinander, gegeben durch den Abstand der Klassenmittelpunkte, beschreibt die Interklassenkovarianzmatrix \mathbf{B} (mit $\bar{\mathbf{m}}$, dem Mittelpunkt aller Klassen):

$$\mathbf{B} = \mathbf{E}\{(\mathbf{m}_{\xi_i} - \bar{\mathbf{m}})(\mathbf{m}_{\xi_i} - \bar{\mathbf{m}})^T\}. \quad (8.5)$$

Die Ausdehnung der Klassen wird als die mittlere Kovarianzmatrix der Klassen $\mathbf{W} = \mathbf{E}\{\mathbf{C}_{\xi_i}\}$ definiert. Daraus ergibt sich das Spurkriterium aus dem Quotienten

$$F = \frac{\text{tr } \mathbf{B}}{\text{tr } \mathbf{W}}. \quad (8.6)$$

Weiterhin wird die Qualität über die Klassifikationsrate eines Mahalanobis-Abstandsklassifikators gemessen. Um die Eigenschaften des Klassifikators möglichst wenig zu berücksichtigen, wird dieser sehr einfache Klassifikator gewählt. Er wurde bereits in Kap. 7.2.2 mit den Gleichungen (7.4) und (7.5) beschrieben.

Bestimmung der Trennbarkeit für kleine Datensätze

Für die Untersuchung standen zehn Kleinmotoren (Fensterhebermotoren) zur Verfügung, die den folgenden, bereits in Kap. 2 beschriebenen Fehlerklassen angehören:

Anzahl	Fehlerklasse	Beschreibung
4	A	fehlerfreie Motoren
0	B	allgemeines Geräusch
1	C	Unwucht
6	D	Bürstengeräusche im Kollektor
1	E	Getriebegeräusche

Für die ungestörten Motorsignale wurde zunächst ein Merkmalsdatensatz gefunden, der eine ausreichend gute Klassifikation ermöglicht. Dazu stand ein Merkmalskatalog von 69 Merkmalen aus den folgenden Gruppen zur Verfügung: physikalische Kennwerte (z.B. Crestfaktor), statistische Merkmale (z.B. Kurtosis, Medianwert), Merkmale aus dem Ordnungs- sowie dem Leistungsdichtespektrum (z.B. Frequenz der größten Spektrallinie) und akustischen Pegel (z.B. dB(A)-Wert).

Die Stichprobe ist zu gering, um eine verlässliche statistische Analyse der Grundgesamtheit aller Motoren der Fehlerklassen zu gewährleisten. Da der Schwerpunkt jedoch nicht auf einer präzisen Klassifikation, sondern nur auf der Veränderung der Klassifizierbarkeit durch die Störung bzw. die Signalverarbeitung liegen soll, kann man auch aus dieser Stichprobe wertvolle Informationen ziehen.

Da die Stichprobengröße noch immer zu gering ist, um die Klassenkovarianzen statistisch zu schätzen, soll eine alternative geometrische Methode verwendet werden, die von Kunter (2002) vorgeschlagen wurde (siehe auch Knaak *et al.* (2002)). Bei normalverteilten Klassen liegen die Punkte gleicher Wahrscheinlichkeit auf Ellipsoiden oder bei zwei Merkmalen auf Ellipsen, die Konzentrationellipsoide bzw. -ellipsen genannt werden und deren Größe sich aus der Kovarianzmatrix ergibt. Die zur Wahrscheinlichkeit 0,68 (1σ -Abstand) gehörenden Kurve/Fläche wird durch die quadratische Form

$$(\mathbf{f} - \mathbf{m}_{\xi_i})^T \mathbf{C}_{\xi_i}^{-1} (\mathbf{f} - \mathbf{m}_{\xi_i}) = \text{const}$$

bestimmt. Mit der Eigenwertzerlegung von $\mathbf{C}_{\xi_i} = \mathbf{U}_{\xi_i} \lambda_{\xi_i} \mathbf{U}_{\xi_i}^T$ ergibt sich

$$(\mathbf{f} - \mathbf{m}_{\xi_i})^T \mathbf{U}_{\xi_i} \lambda_{\xi_i}^{-1} \mathbf{U}_{\xi_i}^T (\mathbf{f} - \mathbf{m}_{\xi_i}) = \text{const.}$$

Da \mathbf{C}_{ξ_i} eine symmetrische Matrix ist, ist die Matrix der Eigenvektoren eine Orthogonalmatrix und beschreibt die Drehung des Ellipsoids um ϕ_{ξ_i} im Merkmalsraum infolge der Korrelation zwischen den Merkmalen (bei unkorrelierten Merkmalen liegen die Hauptachsen des Ellipsoids entlang der Achsen des Koordinatensystems, d.h. der Merkmale). Im Fall von zwei Merkmalen ergibt sich damit folgende quadratische Norm (unter Auslassung der Klassenbezeichnung):

$$(\mathbf{f} - \mathbf{m})^T \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{bmatrix} \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} (\mathbf{f} - \mathbf{m}) = \text{const}$$

und aufgelöst

$$\frac{(f_1 - m_1) \cos \phi - (f_2 - m_2) \sin \phi}{\lambda_1} + \frac{(f_1 - m_1) \cos \phi + (f_2 - m_2) \sin \phi}{\lambda_2} = \text{const.} \quad (8.7)$$

Wie man leicht erkennt, hat (8.7) die Form einer gedrehten Ellipsengleichung mit den Halbachsen $a^2 = \lambda_1$ und $b^2 = \lambda_2$ (vgl. Abb. 8.6).

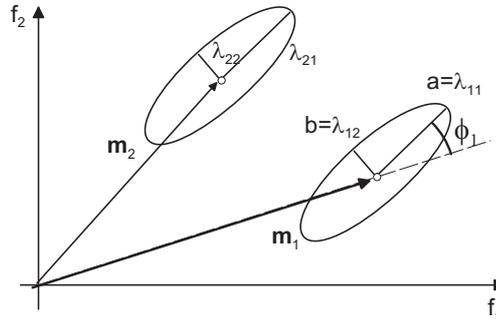


Abb. 8.6: Darstellung der Konzentrationsellipsen zweier Klassen zur geometrischen Bestimmung der Klassenparameter

Damit kann man das Spurkriterium neu definieren, indem man ausnutzt, daß die Spur einer Matrix der Summe der Eigenwerte entspricht. Für die Intraklassenkovarianz ergibt sich allgemein

$$\text{tr } \mathbf{W} = \text{tr} \frac{1}{K} \sum_{k=1}^K \mathbf{C}_{\xi_k} = \frac{1}{K} \sum_{k=1}^K \sum \lambda_{\xi_k, n}$$

und für eine binäre Klassifikation

$$\text{tr } \mathbf{W} = (a_1^2 + b_1^2 + a_2^2 + b_2^2).$$

Für Interklassenkovarianz gilt im binären Fall mit $\text{tr}(\mathbf{xx}^T) = |\mathbf{f}|^2$ und $\bar{\mathbf{m}} = (\mathbf{m}_1 + \mathbf{m}_2)/2$

$$\text{tr } \mathbf{B} = \frac{1}{2} |\mathbf{m}_1 - \bar{\mathbf{m}}|^2 + \frac{1}{2} |\mathbf{m}_2 - \bar{\mathbf{m}}|^2 = \frac{1}{2} |\mathbf{m}_1 - \mathbf{m}_2|^2.$$

Die Werte a und b sowie die Klassenmittelpunkte braucht man nicht mehr statistisch zu schätzen, sondern kann sie aus der Geometrie im Merkmalsraum ermitteln, wie in Abb. 8.7 erkennbar. Zur Bestimmung der großen Halbachse wird der Abstand zwischen dem Klassenmittelpunkt und dem von ihm am weitesten entfernten Vektor dieser Klasse genutzt (mit K , der Menge der Vektoren der Klasse und Auslassung der Klassenbezeichnung).

$$\mathbf{f}_a = \arg \max_{\mathbf{f} \in K} |\mathbf{f}_k - \mathbf{m}| \quad (8.8)$$

Somit ist $a = |\mathbf{f}_a - \mathbf{m}|$ und

$$\tan \phi = \frac{(\mathbf{f} - \mathbf{m}) \cdot [1 \ 0]^T}{(\mathbf{f} - \mathbf{m}) \cdot [0 \ 1]^T}.$$

Die kleine Halbachse ergibt sich aus dem Abstand zwischen der durch \mathbf{f}_a und \mathbf{m} verlaufenden Gerade und dem Vektor \mathbf{f}_b , der am weitesten von ihr entfernt ist.

$$\mathbf{f}_b = \arg \max_{k \neq a \in K} \left(|\mathbf{f}_k - \mathbf{m}|^2 - \left| (\mathbf{f}_k - \mathbf{m})^T \frac{\mathbf{f}_a - \mathbf{m}}{|\mathbf{f}_a - \mathbf{m}|} \right|^2 \right) \quad (8.9)$$

Das Argument der Maximumfunktion in (8.9) stellt die Abstandsberechnung nach Pythagoras dar, wobei der zweite Teil die Projektion von $\mathbf{f} - \mathbf{m}_k$ auf den Vektor $\mathbf{f}_a - \mathbf{m}_k$ angibt, siehe Abb. 8.7. Der Halbachsenabstand b ergibt sich dabei aus (8.7)

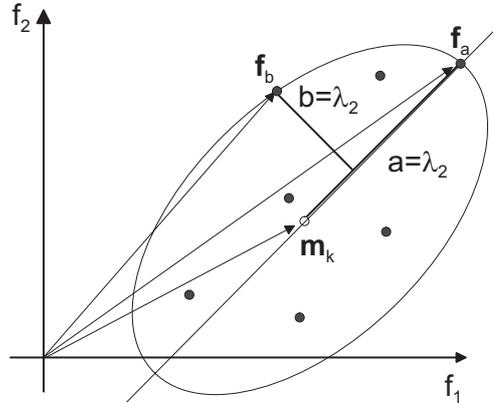


Abb. 8.7: Bestimmung der Ellipsenparameter

$$b^2 = a^2 \frac{[(f_1 - m_1) \sin \phi + (f_2 - m_2) \cos \phi]^2}{a_2 - [(f_1 - m_1) \cos \phi - (f_2 - \mu_2) \sin \phi]^2}. \quad (8.10)$$

Nach dieser Definition liegen alle Vektoren der Stichprobe innerhalb der Ellipse. Sie ist damit größer als die reale 1σ -Ellipse der der Klasse zugrunde liegenden Normalverteilung. Die geometrisch ermittelte Streuung der einzelnen Klassen ist etwas zu groß und damit das Spurkriterium zu klein. Damit wird dem kleinen Stichprobenumfang Rechnung getragen. Außerdem ist nicht der Absolutwert des Spurkriteriums, sondern seine Veränderung infolge der Signalverarbeitung von Interesse. Bei dem so definierten Spurkriterium ist eine hinreichend gute Klassifikation in der Regel möglich, wenn das Spurkriterium größer als eins ist.

Aus den geometrisch ermittelten Parametern kann man nun auch den Mahalanobis-Abstand berechnen und für diesen eine Klassifikationsrate angeben. Der Mahalanobis-Abstand $\varrho_j(\mathbf{f})$ eines Vektors \mathbf{f} zum Klassenmittelpunkt der Klasse i ergibt sich aus der allgemeinen Definition in (7.4)

$$\varrho_j(\mathbf{f}) = \left| \left[\begin{array}{cc} \frac{1}{a_j} & 0 \\ 0 & \frac{1}{b_j} \end{array} \right] \left[\begin{array}{cc} \cos \phi_j & -\sin \phi_j \\ \sin \phi_j & \cos \phi_j \end{array} \right] (\mathbf{f} - \mathbf{m}_j) \right|^2. \quad (8.11)$$

Da die Datenbasis sehr gering ist und die Auflösung der Klassifikationsrate dadurch recht schlecht ist, ist dieses Maß nur bedingt aussagekräftig. Außerdem ist selbst bei hohen Reklassifikationsraten die Generalisierungseigenschaft für unbekannte Vektoren nicht gesichert. Darum wird ein anderer Weg gegangen: Zu einer Fehlklassifikation kommt es, wenn ein Vektor einen kleineren Mahalanobis-Abstand zu einer anderen Klasse als seiner eigenen hat: $\varrho_j(\mathbf{f}_{\xi_i}) < \varrho_i(\mathbf{f}_{\xi_i})$. Äquivalent heißt das, daß der „minimale Mahalanobis-Quotient“ q_{\min}

$$q_{\min}(\xi_i) = \min_{\mathbf{f} \in \xi_i} \frac{\varrho_j(\mathbf{f}_{\xi_i})}{\varrho_i(\mathbf{f}_{\xi_i})} > 1$$

gelten muß, wenn eine hundertprozentige Reklassifikationsrate erreicht werden soll. Kommen unbekannte Vektoren hinzu, ist die Wahrscheinlichkeit, daß sie nicht falsch klassifiziert werden umso größer, je größer q_{\min} ist.⁶ Damit gibt die q_{\min} die Verlässlichkeit des Klassifikators an.

In der vorliegenden binären Klassifikation kann man für die Klasse A (GUT-Motoren) und die Klasse D (Bürstengeräusche) jeweils einen minimalen Mahalanobis-Quotienten angeben. Hier wird nur der minimale Mahalanobis-Quotient der Klasse D untersucht, weil für diese Klasse die Kosten einer Fehlklassifikation (d.h. ein schlechter Motor wird den guten Motoren zugeordnet) am höchsten sind. Die Maßzahl ist wie folgt zu interpretieren: Ist $q_{\min}(D) > 1$ bedeutet das eine Reklassifikationsrate von 100%. Ein höherer Wert gibt eine sichere Klassifikation unbekannter Muster an, ein kleinerer Wert ($q_{\min}(D) < 1$) eine Reklassifikationsrate kleiner 100%. Die Reklassifikationsrate wird dann mit angegeben.

8.3 Ergebnisse der ICA-Verfahren

8.3.1 Auswahl der Verfahren

Für die Untersuchung wurden die Algorithmen in Tab. 8.2 ausgewählt, welche in Voruntersuchungen die besten Resultate hatten. Aus der Gruppe der Beamformer wurde der MV-Beamformer (MVu, MVg) ausgewählt, während auf den in der Sprachverarbeitung gebräuchlicheren ML-Beamformer verzichtet wurde. Denn, wie in Kap. 4.3.2 dargestellt, benötigt letzterer eine Schätzung der Kovarianzmatrix der Nebengeräusche, die in der Sprachverarbeitung in den häufigen Sprachpausen geschätzt werden kann. Das Motorsignal ist hingegen kontinuierlich vorhanden, so daß die Schätzung nur vor dem Ein- oder nach dem Ausschalten erfolgen kann. Durch die relativ lange Zeit für den Hoch- bzw. Auslauf der Maschinen, müßten die Nebengeräusche langzeitstationär sein, wenn eine vor oder nach dem Lauf der Maschine geschätzte Kovarianzmatrix identisch mit Kovarianzmatrix zum Zeitpunkt der eigentlichen Signalaufnahme sein soll. Das kann aber nicht garantiert werden, weil zu viele instationäre Komponenten im Störgeräusch enthalten sind. Die Beamformer funktionieren grundsätzlich mit beliebig vielen Mikrofonen und können damit ebenso viele Störgeräusche unterdrücken. In den folgenden Untersuchungen werden sie mit zwei bzw. drei Mikrofonen untersucht, mit denen das durch eine bzw. zwei Störer verfälschte Motorsignal aufgenommen wird.

Als Repräsentanten der (blinden) ICA-Algorithmen wurden der JADE-, der Infomax- und der FastICA-Algorithmus (JADEa, INFOa und FICAa) jeweils im Frequenzbereich verwendet. Für die Lösung des Permutationsproblem wurden die richtigen Kanäle mit Hilfe der Beampattern oder nach der Stärke der periodischen Spektralmodulation ausgewählt.

Die Beampattern-Methode ist derzeit die beste der allgemeingültigen Methoden, weil sie das Permutationsproblem relativ robust und schnell in den meisten

⁶Das gilt natürlich nur, wenn die Annahme, daß die Klasse normalverteilt ist, zutrifft und die Stichprobe repräsentativ ist.

Gruppe	Algorithmus	Kürzel	Kap.	Ein Stö- rer	Zwei Stö- rer
Beamformer	Mimimum Variance Beamformer mit korrektem Verzögerungsvektor	MVBg	4.3	x	x
	Mimimum Variance Beamformer mit ungenauem Verzögerungsvektor	MVBu	4.3	x	x
(Blinde) ICA, allgemein	FastICA im Frequenzbereich, Lösung des Permutationsproblems über den Vergleich der Arraypattern	FICAa	5.2.1	x	
	Infomax im Frequenzbereich, Lösung des Permutationsproblems nach über den Vergleich der Arraypattern	INFOa	5.2.2	x	
(Blinde) ICA für Motoren	Infomax im Frequenzbereich, Lösung des Permutationsproblems nach Kap. 5.4.2	INFOm	5.2.2	x	x
Constrained ICA	ICA mit geometrischen Randbedingungen mit korrektem Verzögerungsvektor	CICA _g	6.2	x	x
	ICA mit geometrischen Randbedingungen mit ungenauem Verzögerungsvektor	CICA _u	6.2	x	x

Tabelle 8.2: Untersuchte Algorithmen mit ihren Testszenarien

Frequenzbändern löst. Außerdem ist sie nicht auf eine bestimmte Länge der Raumimpulsantwort, wie z.B. der Algorithmus von Parra und Spence (2000), begrenzt. Lediglich bei den sehr tiefen Frequenzen ist der Phasenunterschied sehr gering, so daß sich Fehler durch Rundung oder Rauschen verstärkt bemerkbar machen. Darum wurde in diesem Bereich die Korrelation zwischen den spektralen Komponenten verschiedener Frequenzen des gleichen Ausgangssignals maximiert. Da die Lösungen für das Permutationsproblem derzeit auf zwei Kanäle begrenzt sind, werden die Verfahren auch nur mit einem Störsignal untersucht.

Die Zuordnung nach der Stärke der periodischen Spektralmodulation ist in ihren Einsatzgebiet auf die Rekonstruktion periodischer Signale begrenzt. Dafür erlaubt sie, eine genügend hohe Mikrofonanzahl vorausgesetzt, auch dann die Rekonstruktion dieses Signals, wenn es von mehr als einem Störsignal überlagert wurde, was bei den reinen BSS-Algorithmen nicht möglich ist. Darum wurde es mit dem Infomax-Algorithmus (INFOm) auch für die Störung des Motorsignals durch zwei Störsignale getestet.

Letztlich wurde das ICA-Verfahren mit geometrischer Randbedingung (CICA)

eingesetzt. Da auch bei diesem Verfahren die sehr tiefen Frequenzen durch den geringen Phasenunterschied fehleranfällig sind, wurde für diese Frequenzen FastICA mit der Korrelationsmaximierung eingesetzt. Da die CICA ebenfalls nicht auf zwei Mikrofone begrenzt ist, wurde ihre Performance in beiden Testszenarien (ein bzw. zwei Störsignale) überprüft.

Die Verfahren CICA und MVB benötigen den Einfallswinkel (DOA) oder die Position des Nutzsignals als Voraussetzung. Um den Einfluß von DOA-Schätzfehlern zu untersuchen und die Vergleichbarkeit mit anderen Algorithmen zu verbessern, die nicht auf diese Information angewiesen sind, werden zwei Fälle unterschieden: Verwendung der korrekten Position des Nutzsignals (Zusatz „g“) und eine Position, die um 10° (horizontal) verändert war (Zusatz „u“). Obwohl der Einfallswinkel bei der Maschinendiagnose eigentlich bekannt ist, da man die Position des Maschine und der Arrays recht genau bestimmen kann, wird diese Unterscheidung getroffen, weil die Maschine räumlich ausgedehnt sein und das Nutzsignal aus mehreren Richtungen kommen kann. Außerdem ist die Empfindlichkeit auf Schätzfehler des Einfallswinkels für andere Bereiche von großem Interesse, wenn dieser aus dem Signal bestimmt werden muß.

8.3.2 Quantitativer Vergleich

Ergebnisse in der Testumgebung NST (hallfrei)

Die Ergebnisse mit der Impulsdatenbank NST repräsentieren den nahezu hallfreien Fall. Für ihn sind die Ergebnisse der Störunterdrückung in Abb. 8.8 dargestellt. Nahezu alle Algorithmen zeigen eine gute Störunterdrückung. Lediglich der MV Beamformer kann die Störgeräusche nicht verlässlich unterdrücken, wenn die Einfallrichtung nicht genau geschätzt wird. Dies zeigt die große Abhängigkeit des Algorithmus von der Genauigkeit seiner Parameter, wie bereits theoretisch hergeleitet wurde. Die CICA ist hingegen tolerant gegenüber Parameterungenauigkeiten: Obwohl sich die reale Einfallrichtung um 20° von der angenommenen unterscheidet, wird die Störunterdrückung nicht beeinträchtigt.

Der frequenz aufgelöste SIR für die CICA ist in Abb. 8.9 dargestellt. Nur im sehr niedrigen Frequenzbereich tritt eine erhebliche Dämpfung und sogar ein negativer SIR auf. Dies kann damit begründet werden, daß hier die CICA sondern, sondern die FastICA mit der Korrelationsmethode zur Lösung des Permutationsproblems eingesetzt wurde.

Ergebnisse in der Testumgebung DASBox (150 ms) mit zwei Kanälen

Für einen Raum mit Nachhall sind in Abb. 8.10 Ausschnitte aus den Leistungsdichtespektren gezeigt. In Abb. 8.10a erkennt man, daß das originale Motorsignal eine annähernd periodische Struktur hat, da Frequenzlinien im Abstand der Drehfrequenz auftreten. Im gestörten Signal (Abb. 8.10b) ist diese Struktur in dem dargestellten Frequenzbereich nicht mehr zu sehen. In Abb. 8.10c hingegen, in der das

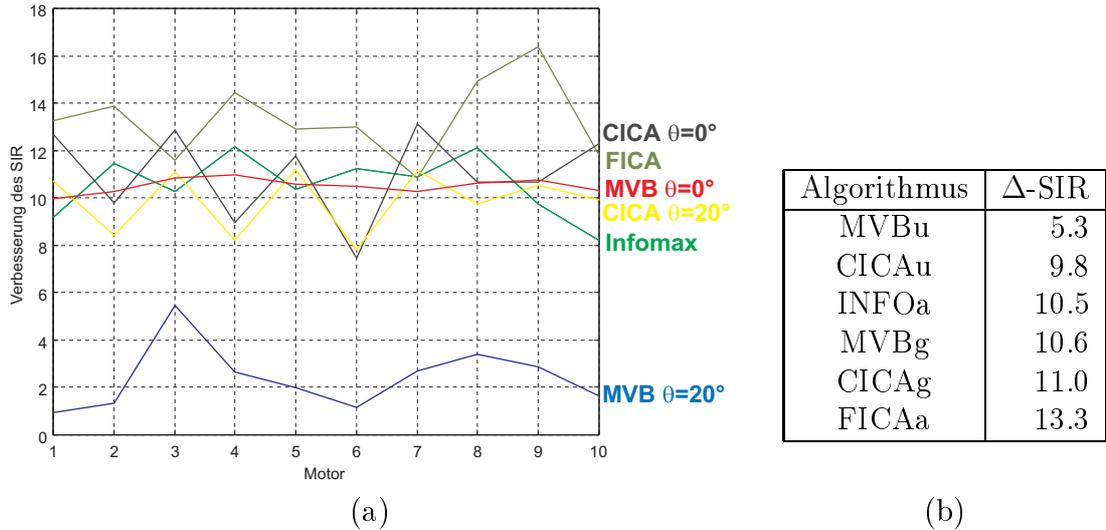


Abb. 8.8: SIR-Verbesserung aller Motoren für $N = M = 2$, (a) einzeln, (b) durchschnittlich

rekonstruierte Signal gezeigt wird, sind die charakteristischen Frequenzlinien mit ihren korrekten Amplituden wieder erkennbar. Damit können bei den rekonstruierten Signalen spektrale Merkmale für die Klassifikation eingesetzt werden.

Die robuste Klassifizierbarkeit ist in Abb. 8.11 und Abb. 8.12 visualisiert. Abb. 8.11a zeigt den Merkmalsraum zweier Merkmale, mit denen die Motoren gut klassifiziert werden können, weil die Klassen klar getrennte Cluster bilden. Die Merkmale sind das Potenzmoment vierter Ordnung (vgl. Definition der Kurtosis in Kap. 5.2.1) und die Lage der zweitgrößten Frequenzlinie. Während die größte Frequenzlinie bei allen Motoren am gleichen Ort liegt, variiert die Lage der zweitgrößten in Abhängigkeit vom Fehlerzustand. Diese Merkmalskombination wurde ausgewählt, weil sie von dem im Kap. 8.2.2 beschriebenen Merkmalskatalog das höchste Spurkriterium hatte.

Bei Auftreten einer Störung verändert sich die Lage der Merkmalsvektoren erheblich. Eine Klassifikation mit diesen Merkmalskombination ist nun unzuverlässig, wie Abb. 8.11b deutlich zeigt: Die beiden Klassen bilden keine getrennten Cluster.

Mit der Signalrekonstruktion werden die ursprünglichen Verhältnisse wieder hergestellt. In Abb. 8.12 ist zu sehen, daß die Cluster nach der Rekonstruktion klar getrennt sind. Das zeigt sich auch quantitativ an den Spurkriterien und Reklassifikationsraten in Tab. 8.3.

Die Tab. 8.3 faßt die Mahalanobis-Abstände und Spurkriterien für die gestörten und mit den einzelnen Algorithmen bearbeiteten Signale zusammen. Die Spurkriterien der ICA-Verfahren liegen durchgängig bei einem Wert um 1,5. Die Mahalanobis-Abstände sind groß genug, um eine Reklassifikationsrate von 100% zu erreichen. Die Ausgangssignale der adaptiven Beamformer sind hingegen nicht für die Klassifikation geeignet, da ihre Mahalanobis-Abstände kleiner eins sind.

Dieses Ergebnis wird auch durch die Verbesserung des SIR bestätigt (vgl. Abb. 8.13). Die besten Resultate werden mit Fast-ICA erzielt. Während hier die

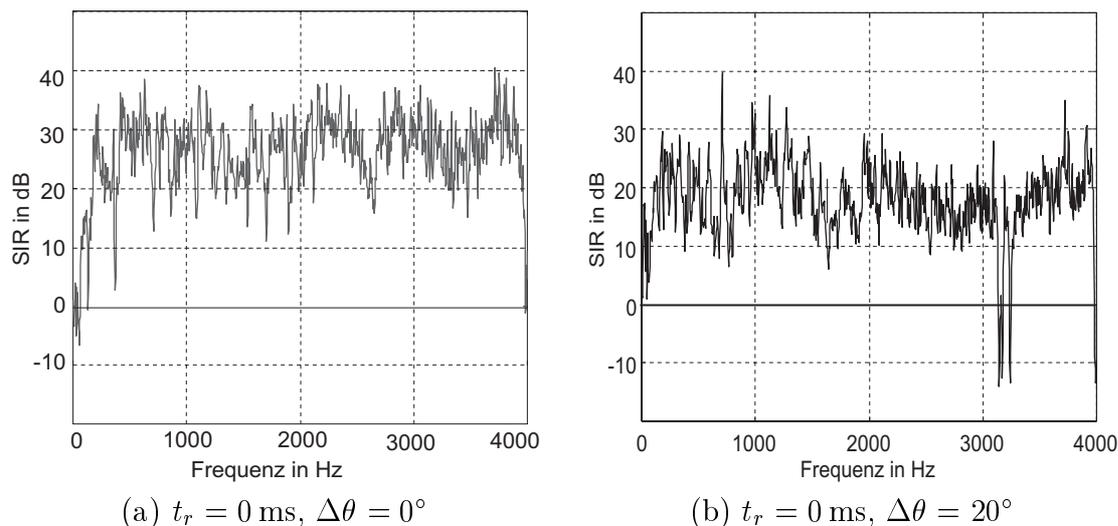


Abb. 8.9: Frequenz aufgelöster SIR für die CICA im Testraum NST mit (a) einem genauen und (b) einem um 20° falsch geschätzten Einfallswinkel

Algorithmus	Spur.	Abstand
Ungestört	3.9	5.1
Gestört	0.2	0.1
FICA	2.0	3.4
INFO	1.1	3.0
CICA _g	1.3	3.1
CICA _u	1.0	1.2
BVB _u	0.7	0.5
MVB _g	0.4	0.4

Tabelle 8.3: Vergleich der (geometrischen) Spurkriterien und der Mahalanobis-Abstände

Verbesserungen des SIR bei ca. 13 dB und bei den anderen ICA-Verfahren bei 10 dB liegen, liegen sie hingegen beim MV-Beamformer nur bei 3 dB. Die Grafik zeigt weiterhin die Toleranz der CICA gegenüber den Parameterschwankungen. Obwohl die reale Einfallrichtung sich um 20° von der realen unterscheidet, ist die Störunterdrückung nicht beeinträchtigt. Dies zeigt die Überlegenheit der ICA-Verfahren im Vergleich zu den MV-Beamformern.

Ergebnisse in der Testumgebung DASBox (150 ms) mit drei Kanälen

Werden die Motorsignale durch mehr als eine gerichtete Nebengeräuschquelle gestört, sind die allgemeinen blinden ICA-Verfahren nicht einsetzbar, weil die allgemeinen Lösungsansätze für das Permutationsproblem auf zwei Observationskanäle beschränkt sind. Darum wird hier die in Kap. 5.4.2 hergeleitete, spezielle Lösung des Permutationsproblems für Maschinensignale genutzt (INFO_m). Im folgenden werden INFO_m, die CICA und der MV-Beamformer verglichen.

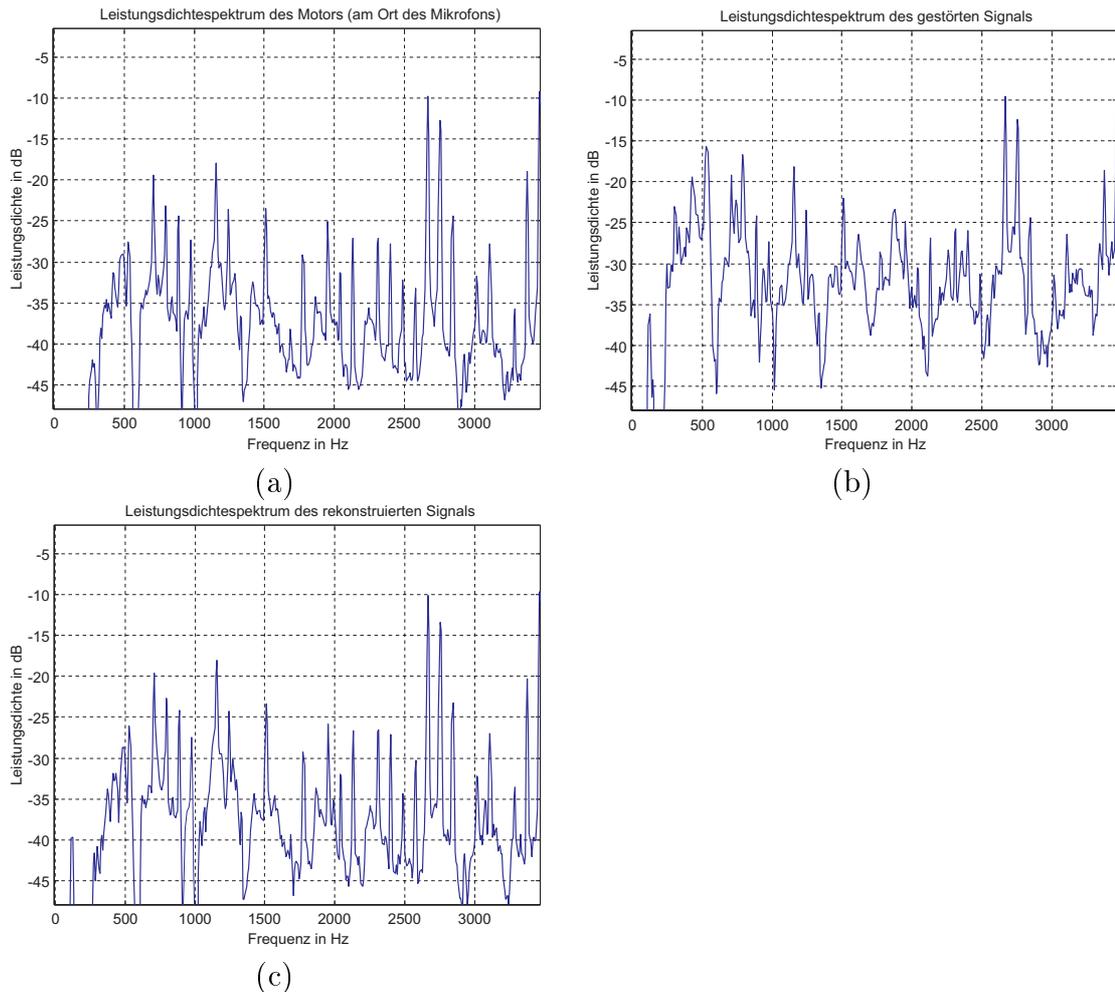


Abb. 8.10: Nahezu perfekte Rekonstruktion der Originalsignals durch den FastICA-Algorithmus. (a) Originalsignal am Ort des Mikrofonarrays, (b) Rekonstruiertes Signal und (c) Gestörtes Signal

Der frequenz aufgelöste SIR ist für die CICA in Abb. 8.14 illustriert. Er liegt in fast allen Frequenzbereichen zwischen 10 und 20 dB. Von singulären Ausreißern abgesehen, kann man erkennen, daß der Algorithmus in jedem Frequenzbereich die Signale trennt und das Permutationsproblem löst.

Der Vergleich der Verfahren in Abb. 8.15 entspricht dem zweikanaligen Fall: Die neuen ICA-Verfahren ermöglichen eine hinreichend guten Rekonstruktion der Signale. Die CICA ist unempfindlich gegen Fehlschätzungen des Parametervektors. Der MV-Beamformer führt zu keiner nennenswerten Unterdrückung der Störungen. Seine Ausgangssignale lassen sich nicht klassifizieren.

Damit sind für den allgemeinen Fall nur die neu eingeführten Verfahren in der Lage, die Störgeräusche hinreichend gut zu unterdrücken. Dies zeigt ihre qualitative Überlegenheit gegenüber den aktuell angewendeten mehrkanaligen Verfahren und ist insofern eine deutliche Verbesserung gegenüber dem derzeitigen Stand der Technik.

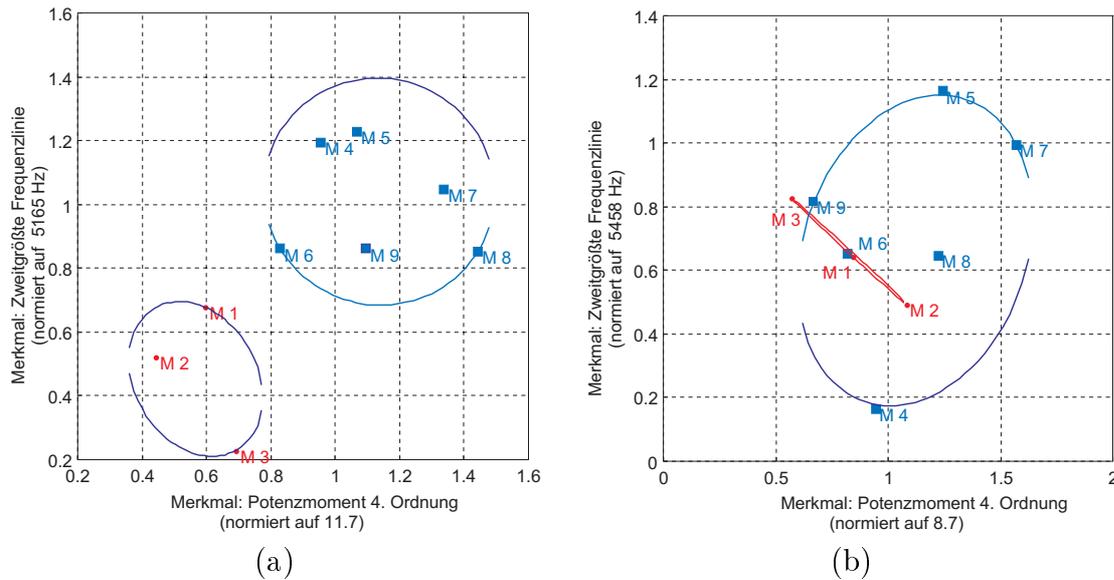


Abb. 8.11: Merkmalsraum für die Klassifikation mit dem Konzentrationsellipsen: (a) Originalsignale, (b) gestörte Signale

8.4 Ergebnisse des Subspace-Algorithmus

Analog zu den zuvor bewerteten Verfahren soll auch der in Kap. 7.2 vorgestellte Algorithmus in Hinblick auf die Klassifikationsrate untersucht werden⁷. Als Meßaufbau wird keine der Testumgebungen, sondern der in Kap. 7.2.3 beschriebene Aufbau direkt in der Fertigungshalle für Waschmaschinen verwendet.

Für die Referenzklassifikation (vgl. Abb. 8.4) wurden die Signale von zwei Testmaschinen (mit und ohne Fehler) mit demselben Mikrofon-Array in einem von der Fertigung abgetrennten, aber nicht völlig schallisolierten Raum aufgenommen. Die zu erkennenden Fehler sind lose Teile, die an die Trommel schlagen. Als Testprogramm wurde darum das Schleudern mit einer Unwucht gewählt. Dabei vibriert die Trommel am meisten, so daß die Schläge klar erkennbar sind. Die Signale wurden während des beginnenden Hochlaufs aufgenommen, weil dabei die lauten Geräusche, die durch die schnelle Rotation entstehen, noch nicht auftreten und die Anschläge des Dämpfers somit deutlicher zu hören sind. Abb. 8.16 zeigt diese Daten.

Zur Merkmalsauswahl kann man erkennen, daß das Signal der fehlerhaften Maschine einen viel höheren Crestfaktor als das der guten Maschine hat. Ein noch sichereres Merkmal, in dem sich das Fehlersignal unterscheidet, ist die Impulshaltigkeit nach DIN/ISO 651 7.3. Ein Impuls ist danach ein plötzliches Ansteigen des Schalldrucks mit mehr als 10 dB/s. Die Impulshaltigkeit ist der Energieanteil solcher Impulse an der Gesamtenergie. Sie garantiert bereits allein eine sichere Unterscheidbarkeit der Klassen, wie Abb. 8.17 zeigt. Darin wird der Merkmalsraum des Merkmals „Impulshaltigkeit“ und eines informationslosen Merkmals für die Visualisierung

⁷Die Veränderung des SIR ist an dieser Stelle nicht sinnvoll, weil das Motorsignal nicht rekonstruiert wird, sondern nur Phasen relativer Ruhe gesucht werden, in den denen Störgeräusche keinen bedeutenden Einfluß auf die Klassifikation haben.

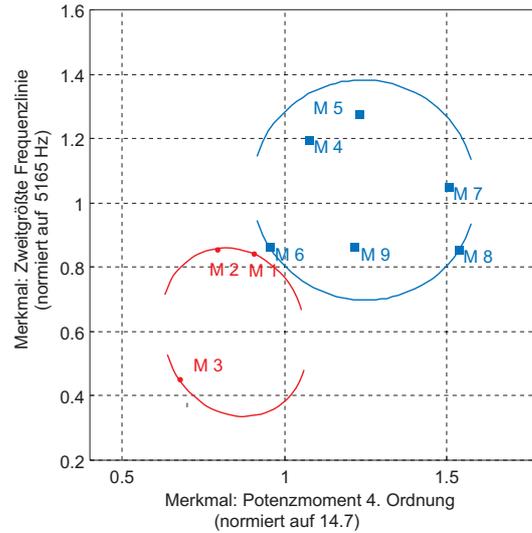


Abb. 8.12: Merkmalsraum für die Klassifikation mit dem Konzentrationsellipsen (mit FastICA rekonstruierte Signale)

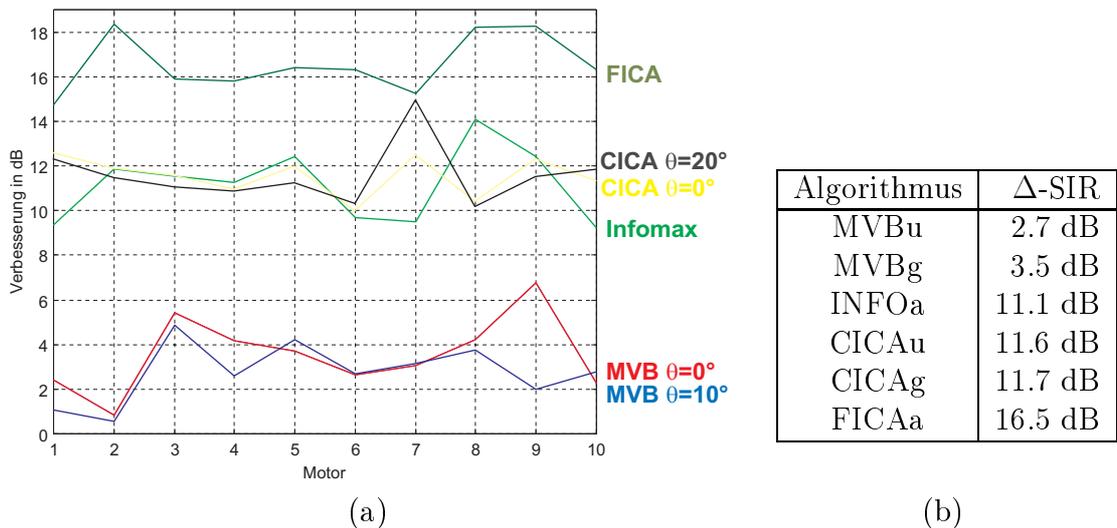
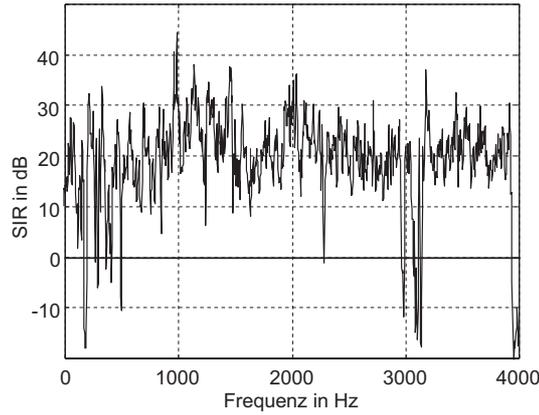


Abb. 8.13: SIR-Verbesserung aller Motoren für $N = M = 2$, (a) einzeln, (b) durchschnittlich

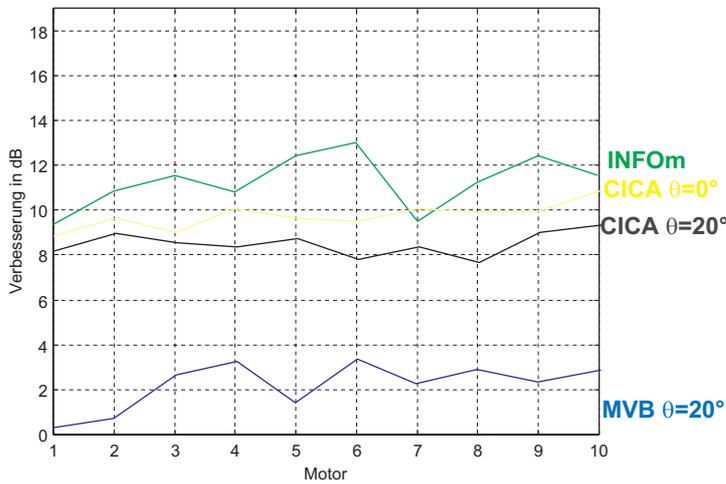
aufgetragen, weil man so die Verteilung der Motoren in den beiden Klassen besser erkennt (für ein Histogramm ist die Besetzung der Klassen zu gering). Das f-Ratio kann nach der statistischen Definition berechnet werden und ist in Tab. 8.4 angegeben. Ebenso sind in diese Tabelle die Klassifikationsraten mit der Impulshaltigkeit und einem einfachen Schwellwertklassifikator aufgeführt.

Zur Bewertung des Algorithmus soll nun die Trennbarkeit der gestörten und der mit dem Subspace-Verfahren bearbeiteten Signale verglichen werden. Die Störung bestand, wie in Kap. 7.2.3 beschrieben, aus einem ebenfalls impulshaltigen Geräusch (Hubtisch). Da nun zu zufälligen Zeitpunkten neue Impulse in das Signalgemisch kommen, steigt die Impulshaltigkeit unabhängig vom Fehler, so daß eine Klassifikation unmöglich wird. Abb. 8.18a/b zeigt in den oberen Kurven die gestör-



(b) $t_r = 150 \text{ ms}$, $\Delta\theta = 20^\circ$

Abb. 8.14: Frequenzaufgelöster SIR für die CICA im Testraum DASBox einem um 20° falsch geschätzten Einfallswinkel



(a)

Algorithmus	Δ -SIR
MVBg	2.2 dB
CICAu	8.5 dB
CICAg	9.7 dB
INFOa	11.2 dB

(b)

Abb. 8.15: SIR-Verbesserung aller Motoren für $N = M = 3$, (a) einzeln, (b) durchschnittlich

Klassifikation	f-Ratio	Fehl- klassifikation: (GUT statt SCHLECHT)	Fehl- klassifikation: (SCHLECHT statt GUT)
Referenzsignale	4.3	0%	0%
gestörte Signale	0.2	45%	27%
bearbeitete Signale	1.2	0%	9%

Tabelle 8.4: f-Ratios und Fehlklassifikationen des Waschmaschinendatensatzes vor und nach der Anwendung des Subspace-Verfahrens

ten Signale, in denen die neuen Impulse deutlich erkennbar sind, und Abb. 8.19a den Merkmalsraum, in dem beide Klassen fast deckungsgleich übereinander liegen.

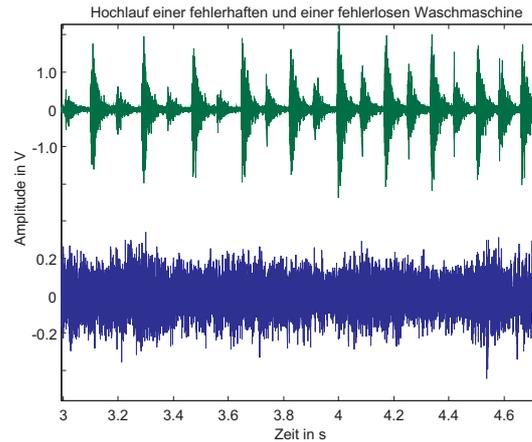


Abb. 8.16: Schallsignal einer Waschmaschine im beginnenden Schleudergang (Hochlauf). Oben sind die Impulse des anschlagenden Reibdämpfers deutlich erkennbar, während er unten fest montiert ist

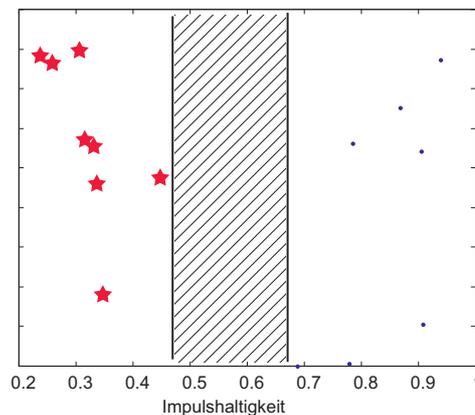


Abb. 8.17: Merkmalsraum der Referenzsignale (Datensatz der Waschmaschinen); das Merkmal auf der y-Achse ist ein informationsloses Merkmal zur Visualisierung.

Das f-Ratio sinkt auf nahezu null.

Nun wird der Subspace-Algorithmus auf die Signale angewendet, der alle Zeiträume ausblendet, in denen mehr als eine gerichtete Quelle aktiv ist, so daß nur die Zeitintervalle in Abb. 8.18 (unten) übrigbleiben. Aus dieser Graphik erkennt man bereits, daß nur die „richtigen“ Impulse im Signal verbleiben. Damit ist die Impulshaltigkeit wieder ein aussagekräftiges Merkmal für eine hinreichend gute Klassifikation (Abb. 8.19b).

Tab. 8.4 faßt die Verbesserung durch das Subspace-Verfahren zusammen. Neben den f-Ratios sind auch die Fehlklassifikationsraten (Off-Diagonalelemente der Verwechslungsmatrix) angegeben, wenn man einen einfachen Schwellwert-Klassifikator einsetzt. Auch mit diesem Maß zeigt sich eine deutliche Verbesserung nach Anwendung des Subspace-Verfahrens.

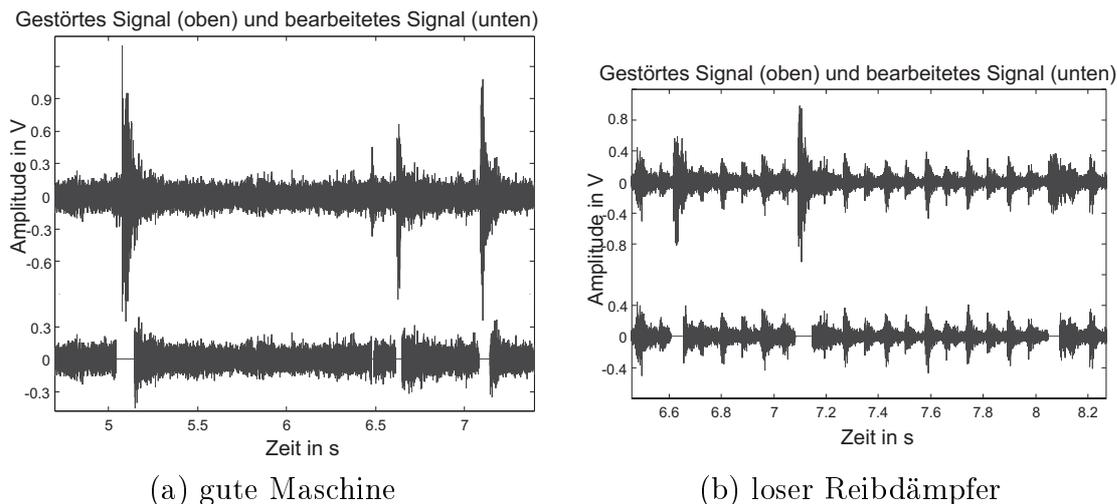


Abb. 8.18: Schallsignale der Waschmaschinen (entsprechend Abb. 8.16) mit der Störung durch einen Hubtisch (oben) und die Zeitintervalle, die für die Klassifikation verwendet werden (unten)

Nicht unerwähnt bleiben sollte auch, daß die Rechenanforderungen des Subspace-Verfahrens gering sind. Die spektrale Schätzung einzelner spektraler Anteile kann mit einem asymmetrischen Bandpaß erfolgen, so daß keine DFT nötig ist. Die Schätzung der Kovarianzmatrix und ihre Eigenwertzerlegung sind ebenfalls sehr schnell, so daß der Algorithmus in Echtzeit arbeiten kann. Die Rechenzeiten sind also so gering, daß die Anforderung, eine Maschine in einem Takt des Fließbandes zu klassifizieren, durch die Vorverarbeitung nicht beeinträchtigt wird.

Zusammenfassend kann man feststellen, daß der Subspace-Algorithmus einsetzbare Signale für eine verlässliche Klassifikation bereitstellt, wenn die Störgeräusche intermittierend sind. Diese Eigenschaft konnte durch Aufnahmen in einer Fabrikhalle, d.h. mit realen Daten in der realen Umgebung für die Maschinendiagnose, verifiziert werden. Damit wurde ein neuer Algorithmus vorgestellt und seine sofortige Anwendbarkeit und sein praktischer Nutzen nachgewiesen.

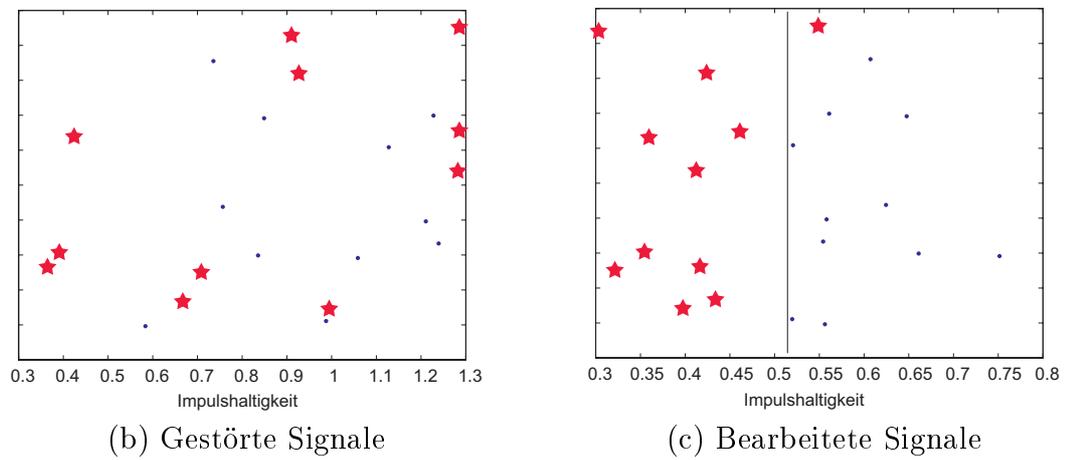


Abb. 8.19: Merkmalsraum der Waschmaschinensignale, das Merkmal auf der y-Achse ist ein informationsloses Merkmal zur Visualisierung.

Kapitel 9

Zusammenfassung

Die Motivation für diese Arbeit war die Unterdrückung von Nebengeräuschen bei der akustischen Güteprüfung, um die aufwendige Dämmung reduzieren zu können. Darum sollte untersucht werden, ob mehrkanalige Verfahren, insbesondere Beamformer und die ICA, für diese Vorverarbeitung eingesetzt werden können.

Da für diese Untersuchung ein tieferes Verständnis der Funktionsweise der mehrkanaligen Verfahren notwendig ist, wurden sie zunächst theoretisch analysiert und verglichen. Dabei wurde als neue Erkenntnis festgestellt, daß die konvolutive ICA äquivalent zu adaptiven Beamformern (ohne a-priori-Wissen über die Einfallrichtungen) ist. Diese Äquivalenz bildet eine systematische theoretische Basis der konvolutiven ICA im Kontext der mehrkanaligen akustischen Signalverarbeitung. Weiterhin wurde sie genutzt, um die Einsatzmöglichkeiten und Grenzen der konvolutiven ICA aufzuzeigen. Nicht zuletzt wurde auf ihrer Basis ein neuer Algorithmus, die „Constrained ICA“ entwickelt, der die Vorteile der beiden Theorien miteinander verbindet.

Dieser Algorithmus wurde sowohl in Mischungen mit Impulsdatenbanken als auch in einem realen Meßraum getestet. Die Signale elektrischer Kleinmotoren, die in diesem Raum mit anderen Signalen gestört wurden, konnten erfolgreich von den Störungen getrennt werden. Das zeigte sich durch einen geringen Anteil der Störungen an der Energie des rekonstruierten Signals und eine hohe Klassifikationsrate mit dem gestörten Signal. Die gute Rekonstruktion konnte auch dann nachgewiesen werden, wenn mehr als zwei Signale abgetrennt werden müssen. Andere ICA-Algorithmen auf dem Stand der Technik sind hingegen auf zwei Signale begrenzt.

Die geringere Parameterempfindlichkeit gegenüber den Beamforming-Algorithmen wurde neben dem Test am praktischen Beispiel auch formal gezeigt: Die analytisch gefaßte Parameterempfindlichkeit des MV-Beamformers sagt aus, daß bereits kleine Abweichungen der angenommenen von den tatsächlichen Parametern die Leistungsfähigkeit der Beamformer sinken läßt. Für die „Constrained ICA“ konnte hingegen nachgewiesen werden, daß sie auch dann das Nutzsignal rekonstruiert, wenn die Parameter von der Realität abweichen, solange die Einfallrichtung nicht eher in die Richtung des Störsignals zeigt.

Da für die Herleitung keine Annahmen über die Art der Signale gemacht wur-

den, ist der Algorithmus für auch andere Bereiche der akustischen Signalverarbeitung nutzbar. Ein Beispiel ist dabei die Sprachverarbeitung, für die auch ungestörte Sprechersignale rekonstruiert werden müssen. Mit Sprachsignalen wurde der neue Algorithmus ebenfalls erfolgreich getestet (vgl. Knaak *et al.* (2003)).

Beschränkt man das Einsatzgebiet auf Maschinengeräusche, kann man deren Eigenschaften, wie die annähernde Periodizität, ausnutzen, um die ICA zu verbessern. Ein solcher Algorithmus wurde ebenfalls entwickelt und erfolgreich getestet.

Schließlich wurde ein Algorithmus für intermittierend auftretende Störungen entwickelt, der unter dieser Annahme sehr robust ist. Er wurde auch direkt an einer Produktionslinie erfolgreich getestet.

Die (konvolutive) ICA in der Akustik steht erst am Anfang ihrer Entwicklung, so daß noch zahlreiche ungelöste Probleme existieren, die bis zu ihrer Anwendung in der Serienprüfung gelöst werden müssen. Dennoch konnte in der vorliegenden Arbeit das große Einsatzpotential der ICA für die Maschinendiagnose gezeigt werden.

Abbildungsverzeichnis

2.1	Struktur eines merkmalsbasierten Fehlerdiagnosesystems	8
2.2	Struktur des vorgeschlagenen Fehlerdiagnosesystems mit einer Erweiterung um einen mehrkanaligen Signalschätzer	10
3.1	„Mehrwegeüberlagerung“ durch Reflexionen an den Wänden	14
3.2	Typische Impulsantwort eines kleinen Raumes mit einer Nachhallzeit von 150 ms	15
3.3	Analogie zwischen Shadingfunktion und einem zeitlichen Fenster	17
3.4	Linienarray: Anordnung (a) und Arraypattern (b)	18
3.5	Arraypattern eines diskreten Linienarrays in Abhängigkeit vom Einfallswinkel mit räumlichem Aliasing	20
4.1	Linienarray: im Fernfeld (a) und Nahfeld (b)	25
4.2	Konventionelle Beamformer: Delay-&Sum-Beamformer (a) und Filter-&Sum-Beamformer (b)	27
4.3	Wellenzahl-Frequenz-Antwort eines DBF	28
4.4	Streudiagramm der Quellensignale	33
4.5	Streudiagramm der aufgenommenen Signale mit den Höhenlinien der Energiefunktion und der Randbedingung.	34
4.6	Definition von α und μ im Streudiagramm	38
4.7	Fehlerwinkel μ (Schätzfehlerwinkel) für die Simulation einer falsch geschätzten Einfallsrichtung	39
4.8	Vergleich der durch grobe Schätzung der Einfallsrichtung und Nachhall entstandenen Schätzfehlerwinkel.	40
4.9	Verstärkung des Nutzsignals (durchgezogen) und des Störsignals (gestrichelt) in Abhängigkeit vom Schätzfehlerwinkel. Die Werte von α betragen 50° und 60°	47
5.1	Grundstruktur eines ICA-System zur Trennung akustischer Mischungen. a) FIR-System. b) rekursives (IIR-) System	53
5.2	Konvergenz eines FastICA-Algorithmus auf dem Einheitskreis	64

5.3	Funktionsverlauf der Verteilungen in (a): p_n^+ (gestrichelt) und p_n^- (gepunktet) sowie der Nichtlinearitäten in (b): g_n^+ (gestrichelt) und g_n^- (gepunktet)	66
5.4	Schematisches Spektrogramm eines Störsignals, bei dem nicht alle Frequenzbänder miteinander korreliert sind.	78
5.5	Frequenzaufgelöster SIR eines FastICA-Algorithmus. Das Permutationsproblem ist in (a) mit dem geometrischen Verfahren gut und mit der Korrelationsmethode (b) schlecht gelöst. Der Gesamt-SIR ist in (b) sehr gering, obwohl in jedem Frequenzbereich eine gute Entmischung erzielt wurde.	79
5.6	Spektrum eines Fensterhebermotors mit den Drehzahlharmonischen (a - Gesamtansicht, b - Ausschnitt)	81
5.7	Bandpaßgefilterte Quellensignale: (a) Störsignal,(b) Motorsignal	83
5.8	Untersuchung von Elektromotoren in einer Produktionshalle	84
5.9	Schätzzraten (a) und -genauigkeiten (b) der Drehzahl von gestörten Motorsignalen	84
6.1	Streudiagramm der Observationssignale mit den Hauptachsen des Konzentrationsellipsoid (Eigenvektoren der Kovarianzmatrix) $\mathbf{u}_1/\mathbf{u}_2$ und dem Steeringvektor des adaptiven Beamformers \mathbf{w}	89
6.2	Streudiagramm der geweißten Signale mit den Vektoren der Entmischungsmatrix $\mathbf{t}_1/\mathbf{t}_2$ und dem transformierten Verzögerungsvektor $\hat{\mathbf{t}}_1$	90
6.3	Streudiagramme der geweißten Signale mit den Randbedingungen von CICA und FastICA	91
6.4	Konvergenz der CICA: Ausschnitt aus dem Streudiagramm der geweißten Signale mit dem Startpunkt der Iteration sowie der Randbedingung. Die Punkte geben die Konvergenz zur richtigen Lösung an.	92
6.5	Energiefunktion auf einer korrekt und fehlerhaft geschätzten Randbedingung mit den Startpunkten	93
6.6	Histogramme für die Störunterdrückung mit einem korrekten Verzögerungsvektor (Simulationen)	93
6.7	Histogramme für die Störunterdrückung mit einem fehlerhaften Verzögerungsvektor $\mu = 12^\circ$ (Simulationen)	94
6.8	Histogramme für die Störunterdrückung für $N = M = 3$ mit einem korrekten Verzögerungsvektor (Simulationen)	94
6.9	Histogramme für die Störunterdrückung für $N = M = 3$ mit einem fehlerhaften Verzögerungsvektor $\mu = 12^\circ$ (Simulationen)	95
6.10	Vektoren im Raum der geweißten Signale	95

6.11	Streudiagramme für \mathbf{x} und \mathbf{z} mit verschiedenen Randbedingungen, die sich aus unterschiedlich genau geschätzten Verzögerungsvektoren ergeben, und den Schätzvektoren des MV-Beamformers und der CICA	97
6.12	Vergleich der Robustheit der CICA (durchgezogene Linie) und des MV-Beamformers (gestrichelte Linie)	98
6.13	Vergleich realer Schätzfehler (durchgezogen) mit den kritischen Winkeln (gestrichelt) für reale Mischungen	98
7.1	Abbildung des untersuchten Motors nach Kunter (2002)	100
7.2	Korrelation zwischen den Störgeräuschen und den Motorsignalen (Kreuzkorrelationskoeffizient)	101
7.3	Schätzung der statistischen Abhängigkeit zwischen den Störgeräuschen und den Motorsignalen	101
7.4	Amplitudenhistogramm einer spektralen Komponente eines Sprachsignals und jeweils eine leptokurtische (a) und eine platykurtische (b) Testverteilung jeweils gleicher Varianz	102
7.5	Amplitudenhistogramm einer spektralen Komponente eines Motorsignals mit einer leptokurtischen (a) und platykurtischen Testverteilung (b) $p(s_{\text{dut}}) = b \cosh cs \cdot e^{-s^2/2}$ mit $c = 1, 8$, alle Verteilungen mit gleicher Varianz	103
7.6	Zwei platykurtische Testverteilungen mit $c = 1, 8$ (durchgezogen) und mit $c = 1$ (gestrichelt), gleiche Varianz	104
7.7	Amplitudenhistogramm einer spektralen Komponente eines Störsignals mit der platykurtischen Testverteilung $p(s_{\text{dut}}) = b \cosh cs \cdot e^{-s^2/2}$ und $c = 2, 5$, alle Verteilungen mit gleicher Varianz	104
7.8	Eigenwertverteilung der Kovarianzmatrix mit einer bzw. zwei aktiven Quellen in dB unter idealen Bedingungen (unkorrelierter diffuser Schall mit deutlich geringer Leistung als die gerichteten Quellen, Punktquellen)	106
7.9	Eigenwertverteilung der Kovarianzmatrix mit einer (a) bzw. zwei (b) aktiven Quellen in dB für eine reale Messung unter den in Kap. 7.2.3 beschriebenen Bedingungen	108
7.10	Streudiagramm der beiden größten Eigenwerte für eine (x) oder zwei (o) aktive Quellen; (a) ideale Simulation und (b) reale Messung	108
7.11	Mekaufbau für die Untersuchung des Subspace-Algorithmus in der Fertigungshalle für Waschmaschinen.	110
8.1	Abmessungen des Mekraumes und Anordnung des Mikrofon-Arrays mit zwei (a) und drei (b) Mikrofonen bzw. Quellen	113
8.2	Beitrag des Nutz- und des Störsignals zum rekonstruierten Signal	116
8.3	Bestimmung der Anteile des Stör- (b) und Nutzsignals (a) am rekonstruierten Signal	116

8.4	Klassifikation an ausgewählten Punkten im Signalfluß von Abb. 8.2	117
8.5	Originalspektrum des Motors (a) und die Verfälschung durch die Filterung mit der Raumimpulsantwort (b): Durchgezogene Linie Signal am Ort des Mikrofons, gestrichelte Linie Übertragungsfunktion der Raumimpulsantwort	118
8.6	Darstellung der Konzentrationsellipsen zweier Klassen zur geometrischen Bestimmung der Klassenparameter	120
8.7	Bestimmung der Ellipsenparameter	121
8.8	SIR-Verbesserung aller Motoren für $N = M = 2$, (a) einzeln, (b) durchschnittlich	125
8.9	Frequenz aufgelöster SIR für die CICA im Testraum NST mit (a) einem genauen und (b) einem um 20° falsch geschätzten Einfallswinkel	126
8.10	Nahezu perfekte Rekonstruktion der Originalsignale durch den FastICA-Algorithmus. (a) Originalsignal am Ort des Mikrofonarrays, (b) Rekonstruiertes Signal und (c) Gestörtes Signal	127
8.11	Merkmalsraum für die Klassifikation mit dem Konzentrationsellipsen: (a) Originalsignale, (b) gestörte Signale	128
8.12	Merkmalsraum für die Klassifikation mit dem Konzentrationsellipsen (mit FastICA rekonstruierte Signale)	129
8.13	SIR-Verbesserung aller Motoren für $N = M = 2$, (a) einzeln, (b) durchschnittlich	129
8.14	Frequenz aufgelöster SIR für die CICA im Testraum DASBox einem um 20° falsch geschätzten Einfallswinkel	130
8.15	SIR-Verbesserung aller Motoren für $N = M = 3$, (a) einzeln, (b) durchschnittlich	130
8.16	Schallsignal einer Waschmaschine im beginnenden Schleudergang (Hochlauf). Oben sind die Impulse des anschlagenden Reibdämpfers deutlich erkennbar, während er unten fest montiert ist	131
8.17	Merkmalsraum der Referenzsignale (Datensatz der Waschmaschinen); das Merkmal auf der y-Achse ist ein informationsloses Merkmal zur Visualisierung.	131
8.18	Schallsignale der Waschmaschinen (entsprechend Abb. 8.16) mit der Störung durch einen Hubtisch (oben) und die Zeitintervalle, die für die Klassifikation verwendet werden (unten)	132
8.19	Merkmalsraum der Waschmaschinensignale, das Merkmal auf der y-Achse ist ein informationsloses Merkmal zur Visualisierung.	133

Tabellenverzeichnis

8.1	Testumgebungen mit den Nachhallzeiten	114
8.2	Untersuchte Algorithmen mit ihren Testszenarien	123
8.3	Vergleich der (geometrischen) Spurkriterien und der Mahalanobis- Abstände	126
8.4	f-Ratios und Fehlklassifikationen des Waschmaschinendatensatzes vor und nach der Anwendung des Subspace-Verfahrens	130

Literaturverzeichnis

- Affes, S. und Y. Grenier (1994). Test of adaptive beamformers for speech acquisition in cars. In: *IEEE Int. Conf. on Signal Proc. Applications and Technology*. pp. 154–159.
- Ahmed, M. S. (1989). Compersion of noisy speech enhancement algorithms in terms of LPC perturbations. *IEEE Transactions on Signal Processing* **37**(1), 121–125.
- Aiordachioaie, D., E. Ceanga, R.-I. Mihalcea und N. Roman (1997). Pre-processing of acoustic signals by neural networks for fault detection and diagnosis of rolling mills. In: *Int. Conf. on Artificial Neural Networks*. pp. 251–256.
- Amari, S. (1997). Neural learning in structured parameter spaces – natural riemannian gradient. *Advances in Neural Information Processing* **9**, 127–133.
- Amari, S., A. Chichocki und H. H. Yang (1996). A new learning algorithm for blind signal separation. *Advances in Neural Information Processing* **8**, 757–763.
- Anemüller, Jörn und Birger Kollmeier (2000). Amplitude modulation for convolutive blind source separation. In: *2nd. Int. Workshop of ICA*. Helsinki.
- Araki, Shoko, Ryo Mukai, Shoji Makino, Tsuyoki Nishikawa und Hiroshi Saruwatari (2003). Fundamental limitation of frequency domain blind source separation for convolved mixture of speech. *IEEE Transactions on Speech and Audio Processing* **11**(2), 109–116.
- Araki, Shoko, Shoji Makino, Robert Aichner, Tsuyoki Nishikawa und Hiroshi Saruwatari (2002). Blind source separation for convolutive mixtures of speech using subband processing. In: *Proc. Workshop on Spectral Methods and Multirate Signal Processing*. Toulouse.
- Araki, Shoko, Shoji Makino, Ryo Mukai und Hiroshi Saruwatari (2001). Equivalence between frequency domain blind source separation and frequency domain adaptive null beamformers. In: *Eurospeech*. pp. 2595–2598.
- Asano, F., Y. Motomura, H. Asoh und T. Matsui (2000). Effect of PCA filter in blind source separation. In: *2nd. Int. Workshop of ICA*. Helsinki.
- Barlow, H.B., T. P. Kaushal und G.L. Mitchison (1989). Finding minimum entropy codes. *Neural Computation*.
- Barschdorff, D. und D. Becker (1991). Der lernendes mustererkennungssystem zur akustischen qualitätsprüfung von lüftermotoren. *at* **39**(1), 46–53.
- Baumann, Wolf, Bert-Uwe Köhler, Dorothea Kolossa und Reinhold Orglmeister (2001). Real time separation of convolutive mixtures. In: *3rd. Int. Conf. on BSS and ICA*. San Diego. pp. 65–69.
- Beilharz, Jörg (1999). Frequenzumrichter gestützte Parameteridentifikation von Asynchronmaschinen kleiner Leistung. Dissertation. TU Berlin. Berlin.

- Beilharz, Jörg, Karsten Röpke und Dieter Filbert (1995). Statistische und neuronale Verfahren zum nicht-überwachten klassifikatorentwurf. *at* **43**, 46–53.
- Bell, Anthony und Terence Sejnowski (1995). An information-maximization approach blind source separation and blind deconvolution. *Neural Computation* **7**(6), 1004–1034.
- Belouchrani, A., K. Aded-Meraim und J. Cardoso (1997). A blind separation technique using second order statistics. *IEEE Transactions on Signal Processing* **45**(2), 434–444.
- Bingham, E. und A. Hyvärinen (2000). A fast fixed point algorithm for independent component analysis of complex valued signals. Technical report. <http://www.cis.hut.fi/projects/ica/>. Helsinki.
- Bitzer, J., K.U. Simmer und K.-D. Kammeyer (1999). Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech signals. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Vol. 5.
- Bitzer, J. und K.U. Simmer (2001). Superdirective microphone arrays. In: *Microphone Arrays* (M. Brandstein und D. Ward, Eds.). Springer. Berlin. pp. 19–37.
- Bouquin-Jeannes, R. Le, A. Azirani und G. Faucon (1997). Enhancement of speech signals degraded by coherent and incoherent noise using a cross-spectral estimator. *IEEE Transactions on Signal Processing* **5**(5), 484–487.
- Brandwood, D.H. (1983). A complex gradient operator and its application in adaptive array theory. *IEE Proceedings* **130**, 11–16.
- Bronstein, Ilja N., Konstantin A. Semendjajew, Gerhard Musiol und Heiner Mühlig (2000). *Taschenbuch der Mathematik*. 5 ed.. Verlag Harry Deutsch. Frankfurt.
- Buckley, K. M. und L. J. Griffith (1986). An adaptive generalized sidelobe canceller with derivative constraints. *IEEE Transactions on Antennas and Propagation* **34**(3), 311–319.
- Capon, J. (1967). High-resolution frequency-wavenumber estimation. *Proceedings of the IEEE* **57**, 1408–1418.
- Cardoso, J.-F. und A. Souloumiac (1996). Jacobi angles for simultaneous diagonalization. *SIAM Journal of Matrix Analysis and Applications* **17**(1), 161–164.
- Cardoso, Jean-Francois (1998). Blind signal separation: statistical principles. *Proceeding of the IEEE* **89**(10), 2009–2025.
- Cardoso, Jean-Francois (1999). High-order contrast for independent component analysis. *Neural Computation* **11**(1), 157–192.
- Chalmers, B.J. et al. (1988). *Electric Motor Handbook*. Butterworths. London.
- Chan, Dominic, Peter Rayner und Simon Godsill (1996). Multi-channel signal separation. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. pp. 649–652.
- Cherry, G. und E. Collin (1953). Some experiments on recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America* **24**, 2608–2612.
- Chichocki, Andrzej und Rolf Unbehauen (1996). Robust neural networks with on-line learning for blind identification and blind separation of sources. *IEEE Transactions on Signal Processing* **43**(11), 894–906.
- Cichocki, A. und S. Amari (2002). *Adaptive Blind Signal and Image Processing*. John Wiley. Chichester UK.

- Cover, T. und Thomas, J., Eds.) (1991). *Elements of information theory*. Vol. 1. Jon Wiley. New York.
- Dapena, Adriana, Monica Bugallo und Luis Castedo (2001). Separation of convolutive mixtures of temporally-withe signals: A novel frequency domain approach. In: *3rd. Int. Conf. on BSS and ICA*. San Diego.
- Davies, M. E. (2001). Audio source separation. In: *Mathematics in Signal Processing V*. Oxford University Press. Oxford.
- Drews, Martin (1997). Mehrkanal-Geräuschreduktion mit adaptiven Wiener-Kolmogorov-Filter.
- Drews, Martin (1999). Mikrofonarrays und mehrkanalige Signalverarbeitung zur Verbesserung gestörter Sprache. Dissertation. TU Berlin. Berlin.
- Ehlers, F. und H. G. Schuster (1997). Blind separation of convolutive mixtures and an application in automotive speech recognition in a noisy environment. *IEEE Transactions on Signal Processing* **45**(10), 2608–2612.
- Elko, G.W. (1996). Microphone array systems for handy-free telecommunication. *EURASIP Speech Communication* **20**, 299–240.
- Fabry, P., Ch. Servièrè und J.L. Lacoume (1998). Improving signal subspace for blind source separation in the context of spatially correlated noises. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Vol. 4. pp. 2377–2380.
- Fabry, P. und Ch. Serviere (2000). Blind separation of noisy harmonic signals using orthogonal techniques for rotating machine diagnosis. In: *2nd. Int. Workshop of ICA*. Helsinki.
- Fausten, Michael (2002). Mustererkennung an Pkw-Startermotoren mit adaptiven Fuzzy-Filtern. Dissertation. TU Berlin. Berlin.
- Feldmann, K. und W. Keiper (1994). Gehörgerechte Analyse technischer Schalle. Bosch Technische Berichte 56. Bosch GmbH.
- Filbert, Dieter (2003). Technische diagnose zur qualitätskontrolle an kleinstmotoren. *at*. in Druck.
- Fischer, R. (1995). *Elektrische Maschinen*. Carl Hanser Verlag. München, Wien.
- Florenco, D., S. Nakamura und K. Shikano (2001). Multichannel filtering for optimum noise reduction in microphone arrays. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Salt Lake City. pp. 197–200.
- Frost, O. L. (1972). An algorithm for linearly constrained adaptive antenna array processing. *Proceedings of the IEEE* **60**(8), 926–935.
- Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. 2 ed.. Academic Press. New York.
- Griffin, M.F., S. Tulpule und A.M. Finn (1991). Signal enhancement for diagnostic signal analysis [rocket engines]. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Vol. 5. pp. 3281–3284.
- Griffith, L. J. und C. W. Jim (1982). An alternative approach to linearly constrained adaptive beamforming. *IEEE Transactions on Antennas and Propagation* **30**(1), 27–34.
- Gühmann, Clemens (1995). Stromanalyse zur Diagnose seriengefertigter Universalmotoren. Dissertation. TU Berlin. Berlin.

- Günther, B. C. et al. (1994). *Technische Akustik - Ausgewählte Kapitel*. 5 ed., expert. Renningen-Malmsheim.
- Haykin, Simon (1994). *Neural Networks*. MacMillan. Englewood Cliffs.
- Haykin, Simon (1996). *Adaptive filter theory*. 3 ed., Prentice Hall, Englewood Cliffs, NJ.
- Haykin, Simon, Ed.) (2000). *Unsupervised adaptive filtering*. Vol. 1. John Wiley. New York.
- Heckl, M. und H. A. Müller (1994). *Taschenbuch der Technischen Akustik*. 2 ed., Springer. Berlin.
- Hoffmann, Rüdiger (1998). *Signalanalyse und -erkennung*. Springer. Berlin.
- Hoshuyama, O. und A. Sugiyama (2001). Robust adaptive beamforming. In: *Microphone Arrays* (M. Brandstein und D. Ward, Eds.). Springer. Berlin. pp. 87–106.
- Hua, Y., J. An und Y. Xian (2001). Blind equalization of FIRMIMO channel by BDS. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Salt Lake City. pp. 2157–2160.
- Hyvärinen, Apo, Juha Karhunen und Erkki Oja (2001). *Independent component analysis*. John Wiley. New York.
- Hyvärinen und E. A., Oja (1999). Independent Component Analysis: A Tutorial. Technical report. <http://www.cis.hut.fi/projects/ica/>. Helsinki.
- ichi Amari, Shun (1997). Geometry of semiparametric models and applications. In: *ISI*. Istanbul.
- Ikram, M.Z. und D.R. Morgan (2000). Exploring permutation inconsistency in blind separation of speech signals. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. pp. 1041–1044.
- Jabloun, F und B. Champagne (2001). A multi-microphone signal subspace approach for speech enhancement. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Salt Lake City. pp. 201–205.
- Jahnke, Ronny (2001). Beitrag zur Akustischen Qualitätsprüfung im Bereich industrieller Fertigung. Dissertation. TU Berlin. Berlin.
- Johnson, Don und Dan Dudgeon (1993). *Array Signal Processing: Concepts and Techniques*. Prentice Hall, Englewood Cliffs, NJ.
- Jones, Douglas (1999). A new method for blind separation nonstationary sources. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Phoenix, USA. pp. 2893–2896.
- Kajala, M. und M. Hämmäläinen (2001). Filter-and-sum beamformer with adjustable filter characteristics. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Salt Lake City. pp. 2917–2920.
- Kammeyer, K. D. und K. Kroschel (1998). *Digitale Signalverarbeitung*. 4 ed., Teubner. Stuttgart.
- Knaak, Mirko, Matthias Kunter und Dieter Filbert (2002). Blind source separation for acoustical machine diagnosis. In: *4th IEEE International Conference on Digital Signal Processing*. Santorini, Greece.
- Knaak, Mirko, Michael Fausten und Dieter Filbert (2001). Acoustical machine monitoring using blind source separation. In: *Proc. of 4th Int. Conf. on acoustical and vibratory surveillance methods and diagnostics techniques*. Compiegne, France. pp. 401–412.
- Knaak, Mirko, Shoko Araki und Shoji Makino (2003). Geometrically Constrained ICA for convolutive mixtures of sound. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Hong Kong.

- Knaak, Mirko und Dieter Filbert (2001a). Acoustical semi-blind source separation for machine monitoring. In: *3rd. Int. Conf. on BSS and ICA*. San Diego. pp. 361–366.
- Knaak, Mirko und Dieter Filbert (2001b). ICA-basierte Signalrekonstruktion mit periodischer Spektralmodulation. In: *XV. Meßtechnisches Symposium des Arbeitskreises der Hochschullehrer der Meßtechnik* (Helmuth Röck, Ed.). Shaker. Aachen. pp. 84–92.
- Knaak, Mirko und Dieter Filbert (2002). Semi-blinde Geräuschrekonstruktion für die technische Diagnose. *tm - Technisches Messen* **69**(4), 169 – 174.
- Knaak, Mirko und Dieter Filbert (2003). Acoustic source number estimation and its application to machine monitoring. In: *IFAC Symp. on Fault Detection, Supervision and Safety of Technical Processes (SAFEPROCESS)*. Washington, D.C.
- Knorr, Sebastian (2002). Implementierung von Verfahren zur blinden Trennung konvolutiver Signalgemische im Frequenzbereich. Diplomarbeit. TU Berlin. Berlin.
- Köhler, Bert-Uwe (2000). Realzeitfähige Blinde Quellentrennung am Beispiel elektroenzephalographischer Daten. Dissertation. TU Berlin.
- Kunter, Matthias (2002). Einsatz der blinden Quellentrennung zur Fehlerklassifikation in Elektromotoren. Diplomarbeit. TU Berlin. Berlin.
- Kurita, S., H. Saruwatari, S. Kajita, T. Takeda und F. Itakura (2000). Evaluation of blind signal separation methods using directivity pattern under reverberant conditions. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Istanbul. pp. 3140–3143.
- Kurpjuhn, T., M. Ivrlag und J. Nossek (2001). Vandermonde invariance transformation. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Salt Lake City. pp. 2929–2933.
- Lambert, Russell (1996). Multichannel Blind Deconvolution: FIR algebra and separation of multipath mixtures. Dissertation. University of Southern California. Los Angeles.
- Lambert, Russell, Marcel Joho und Heinz Mathis (2001). Polynomial singular values for number of wideband sources estimation and principal component analysis. In: *3rd. Int. Conf. on BSS and ICA*. San Diego. pp. 379–383.
- Lee, S.K. und P.R. White (1996). Fault diagnosis of rotating machinery using a two-stage adaptive line enhancer. In: *IEE Colloquium on Modeling and Signal Processing for Fault Diagnosis*. pp. 6/1–6/6.
- Lee, T. W., Anthony Bell und Russell Lambert (1997a). Blind separation of delayed and convolved sources. *Advances in Neural Information Processing* **9**, 758–764.
- Lee, Te-Won (1997). Independent Component Analysis: Theorie und Applications. Dissertation. TU Berlin. Berlin.
- Lee, Te-Won, Anthony Bell und Reinhold Orglmeister (1997b). Blind source separation of real world signals. In: *Int. Conf. on Neural Networks*. Houston. pp. 564–569.
- Lee, T.W. (1998). *Independent component analysis - Theory and applications*. Kluwer.
- Lindgren, U.A. und H.; Broman (1998). Source separation using a criterion based on second-order statistics. *IEEE Transactions on Signal Processing* **46**(7), 1837–1850.
- Lüke, H. D. (1995). *Signalübertragung*. Springer. Berlin.

- Luo, H., R.W. Liu, T. Lin und X. Li (2001). The autocorrelation matching method for distributed MIMO communications over unknown FIR channels. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Salt Lake City. pp. 2161–2164.
- MacKay, David (2001). *Information Theory, Pattern Recognition and Neural Networks*. University of Cambridge. <http://www.inference.phy.cam.ac.uk/mackay/itprnn/book.html>.
- Mansour, Ali und Noboru Ohnishi (1999). Multichannel blind separation of sources algorithm based on cross-cumulant and the Levenberg-Marquard method. *IEEE Transactions on Signal Processing* **47**(11), 3172–3175.
- Mathis, Heinz, Marcel Joho und George Moschytz (2000). A simple threshold nonlinearity for blind separation of sub-gaussian signals. In: *IEEE Int. Symposium on Circuits and Systems*. Vol. 4. Genvea. pp. 489–492.
- Matsouka, K., M. Ohya und M. Kawamoto (1995). A neural net for blind separatoins fo nonstationary signals. *Neural Networks* **8**(3), 411–419.
- McCowan, I. und S. Sridharan (2001). Micropone array subband speech recognition. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Salt Lake City. pp. 185–189.
- Mejuto, Cristina, Adriana Dapena und Luis Castedo (2000). Frequency-domain infomax for blind separation of convolutive mixtures. In: *2nd. Int. Workshop of ICA*. Helsinki.
- Mejuto, Cristina und Jose C. Principe (1999). A second-order method for blind source separation of convolutive mixtures. In: *Int. Workshop of ICA*. Aussois, France. pp. 395–400.
- Mitianoudis, N. und M. Davies (2001). New fixed-point ica algorithms for convolved mixtures. In: *3rd. Int. Conf. on BSS and ICA*. San Diego. pp. 633–638.
- Moczala, H. et al. (1998). *Small Electric Motors*. IEE. London.
- Molgedy, L. und H. G. Schuster (1994). Separation of a mixture of independent signal using time delayed correlations. *Phys. Rev. Lett.* **72**(23), 3634–3637.
- Mukai, Ryo, Shoko Araki und Shoji Makino (2001). Separation and dereverberation performance of frequency domain blind source separation. In: *3rd. Int. Conf. on BSS and ICA*. San Diego.
- Murata, S., S. Ikeda und A. Ziehe (1998). An approach to blind source to blind source separation based on temporal structure of speech signals. Technical Report 2. Riken Brain Science Institute. <http://www.bsp.brain.riken.go.jp>.
- Murata, S., S. Ikeda und A. Ziehe (1999). An approach to blind source separation based on temporal structure of speech signals. *IEEE Transactions on Signal Processing*.
- Murray, A. und J. Penman (1997). Extracting useful higher order features for condition monitoring using artificial neural networks. *IEEE Transactions on Signal Processing* **45**(11), 2821–2828.
- Niemann, H. (1983). *Klassifikation von Mustern*. Springer. Berlin.
- Nikias, Ch. und A. Petropulu (1993). *Higher order spectra analysis*. Prentice Hall. Englewood Cliffs.
- Nishikawa, Tsuyoki, Hiroshi Saruwatari und Kyohiro Shikano (2002). Blind source separation based on multistage ICA using frequency-domain ICA and time-domain ICA. In: *ICFS*. Kyoto.
- Nishura, T., S. Nakamura und K. Shikano (2001). Speech enhancement by multiple beamforming with reflection signal equalization. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Salt Lake City. pp. 189–192.

- Noll, Peter (1974). Adaptive quantizing in speech coding systems. In: *Int. Zurich Seminar on Digital Comm.*, pp. B3.1–B3.6.
- Nunn, D. (1983). Performance assessment of a time domain adaptive antenna processor in a broadband environment. *Proceedings of IEE*.
- O'Donnell, P. (1985). Report of large motor reliability survey of industrial and commercial installations, Part I and II. *IEEE Transactions on Industry Applications* **21**(4), 853–872.
- O'Donnell, P. (1987). Report of large motor reliability survey of industrial and commercial installations, Part III. *IEEE Transactions on Industry Applications* **23**(1), 153–158.
- Oppenheim, A. und R. Schafer (1995). *Zeitdiskrete Signalverarbeitung*. Oldenbourg. München, Wien.
- Parra, Lucas (1999). An introduction to independent component analysis and blind source separation. Vorlesungsskript. Columbia University. <http://humanism.org/lucas/publish/tutorial-bss.pdf>.
- Parra, Lucas, Clay Spence und B. de Vries (1998). Convolutional blind source separation based on multiple deconvolution. In: *IEEE Workshop on Neural Networks for Signal Processing NNSP*. Cambridge, UK.
- Parra, Lucas und Christopher Alvino (2001). Geometric source separation: Merging convolutional source separation with geometric beamforming. In: *IEEE Workshop on Neural Networks for Signal Processing NNSP*. pp. 273–282.
- Parra, Lucas und Clay Spence (2000). Convolutional blind source separation of non-stationary signals. *IEEE Transactions on Speech and Audio Processing* **8**, 320–327.
- Parson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **2**, 559–572.
- Parsons, Thomas (1987). *Voice and Speech Processing*. Mc Graw-Hill. New York.
- Pearlmutter, Barak A. und Lucas C. Parra (1996). A context-sensitive generalization of ica. In: *Int. Conf. on Neural Information Processing*, Hong Kong.
- Pham, Dinh-Tuan (2000). Blind separation of instantaneous mixtures of sources based on order statistics. *IEEE Transactions on Signal Processing* **48**(2), 363–375.
- Pham, Dinh-Tuan und Jean-Francois Cardoso (2000). Blind separation of instantaneous mixtures of non stationary sources. In: *2nd. Int. Workshop of ICA*. Helsinki.
- Rahbar, Kamran und James Reilly (2001). Blind source separation for convolved sources by joint approximate diagonalization of cross density matrices. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Salt Lake City. pp. 2745–2748.
- Rojas, R. (1996). *Theorie der neuronalen Netze: Eine systematische Einführung*. 4 ed.. Springer Verlag. Berlin.
- Röpke, Karsten (1997). Statistische und modellgestützte Verfahren zur Klassenbildung bei der Diagnose von Universalmotoren. Dissertation. TU Berlin. Berlin.
- Ruske, G. (1988). *Automatische Spracherkennung*. Oldenbourg. München, Wien.
- Saruwatari, Hiroshi, S. Kurita und T. Takeda (2001). Blind source separation combining frequency-domain ICA and beamforming. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Salt Lake City. pp. 2733–2736.

- Sawada, H., R. Mukai, Shoko Araki und Shoji Makino (2002a). Polar coordinate based nonlinear function for frequency domain blind source separation. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Vol. 1. Orlando.
- Sawada, Hiroshi, Shoko Araki, Ryo Mukai und Shoji Makino (2002b). Blind source separation with different sensor spacing and filter length for each frequency range. In: *IEEE Workshop on Neural Networks for Signal Processing NNSP*.
- Schobben, Daniel, Kari Torkkola und Paris Smaragdis (1999). Evaluation of blind separation methods. In: *Int. Workshop of ICA*. Aussois, France.
- Schobben, Daniel und Piet Sommen (1998). A new blind signal separation algorithm based on second order statistics. In: *Proc. on Int. Conf. on Signal and Image Processing*. Las Vegas. pp. 564–569.
- Schwarz, Joachim (1988). *Digitale Verarbeitung stochastischer Signale*. Oldenbourg. München, Wien.
- Schwarz, Matthias (2002). Skript zur Vorlesung Statistik I und II. Vorlesungsskript. Universität Mainz, Lehrstuhl Prof. Sauerheimer. <http://wiwi.uni-mainz.de/vwl/sauerheimer/matthias/skript.pdf>.
- Servière, Ch. (1998). Feasibility of source separation in frequency domain. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Vol. 4. pp. 2085–2088.
- Servière, Ch. und P. Fabry (2001). Blind separation of harmonic signals for rotating machine diagnosis. In: *4th International Conference on acoustical and vibratory Surveillance Methodes and Diagnostics Techniques*. Compiegne. pp. 129–138.
- Shutao, Zhang und Ian Li-Jin Thng (2001). Pre-steering derivative constraint for robust broadband antenna arrays. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Salt Lake City. pp. 2913–2916.
- Smaragdis, Paris (1998). Blind separation of convolved mixtures in the frequency domain. In: *Int. Workshop on Independence and Art. Neural Networks*. Teneriffe, Spain.
- Stearns, Samuel D. und Don R. Hush (1994). *Digitale Verarbeitung analoger Signale*. 6 ed.. Oldenbourg. München, Wien.
- Steinberg, Bernard D. (1976). *Principles of Aperture and Array System Design : Including Random and Adaptive Arrays*. John Wiley & Sons.
- Stepper, Kay (1999). Ein Beitrag zur akustischen Güteprüfung von Komponenten der Kfz.- und Automatisierungstechnik. Dissertation. TU Berlin. Berlin.
- Sydow, C. (1996). Selbstausrichtende Mikrofonarraysysteme. Dissertation. TH Darmstadt. Darmstadt.
- Therrien, Charles W. (1992). *Discrete random signals and statistical signal processing*. Prentice Hall, Englewood Cliffs, NJ.
- Tong, L.L. und Y.S. Soon (1991). On the indeterminacy and identifiability of blind identification. *IEEE Trans. Circuits and Systems* **38**, 499–509.
- Torkkola, Kari (1996). Blind separation of convolved sources based on information maximization. In: *IEEE Workshop on Neural Networks for Signal Processing NNSP*. Kyoto.
- Torkkola, Kari (1999). Blind separation for audio signals - are we there yet?. In: *Int. Workshop of ICA*. Aussois, France. pp. 239–245.

- Torkkola, Kari (2000). Blind separation of delayed and convolved mixtures. In: *Unsupervised adaptive filtering* (Simon Haykin, Ed.). Vol. 1. John Wiley. New York.
- van Rooijen, W.F.G., E.Y.L. Ling, Futuoshi Asano und Kiyoshi Yamamoto (2002). Acoustic source number estimation using support vector machine and its application to source localization/separation system. *IEICE EA* **2002-41**, 25–30.
- Vaseghi, Saeed V. (1996). *Advanced Signal Processing and Digital Noise Reduction*. 1 ed.. John Wiley. New York.
- Viberg, Mats (2002). Model-based signal processing. Vorlesungsskript. Calmers University. Sweden. <http://www.s2.chalmers.se/~viberg/MBSP/kay2002.html>.
- Wang, F., V. Balakrishnan, P. Zhou, Chen J, R. Yang und C. Frank (2001). Optimal array pattern synthesis using semidefinite programming. In: *Int. Confer. on Acoustics, Speech, and Signal Proc. ICASSP*. Salt Lake City. pp. 2925–2928.
- Ward, D., R. Kennedy, und R. Williamson (2001). Constant directivity beamformer. In: *Microphone Arrays* (M. Brandstein und D. Ward, Eds.). Springer. Berlin. pp. 3–16.
- Weinstein, Ehud, Meir Feder und Alan V. Oppenheim (1993). Multi-channel signal separation by decorrelation. *IEEE Transactions on Speech and Audio Processing* **1**(4), 405–413.
- Wu, Hsiao-Chun und Jose C. Principe (1999). Simultaneous diagonalization in the frequency domain (sdif) for source separation. In: *Int. Workshop of ICA*. Aussois, France. pp. 245–249.
- Yellin, Daniel und Ehud Weinstein (1994). Criteria for multichannel signal separation. *IEEE Transactions on Signal Processing* **42**(8), 2158–2168.
- Ypma, Alexander und Amir Leshem (2000). Blind separation of machine vibration with bilinear forms. In: *2nd. Int. Workshop of ICA*. Helsinki.
- Zelinski, Rainer (1996). Mikrofon-array mit superdirektiven eigenschaften zur sprachsignalübertragung. *Frequenz* **50**, 198–204.
- Ziehe, Andreas und Klaus-Robert Müller (1998). TDSEP - an efficient algorithm for separation using time structure. In: *Int. Conf. on Artificial Neural Networks ICANN*. pp. 675–680.
- Zurmühl, Rudolf und Sigurd Falk (1992). *Matrizen und ihre Anwendung*. Springer. Berlin.