

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

Neue Ansätze zur Lösung des
Skalierungs- und Permutationsproblems
in konvolutiver blinder Quellentrennung

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck
- Aus der Technisch-Naturwissenschaftlichen Fakultät -

vorgelegt von
Radoslaw Mazur
aus Breslau
Lübeck 2010

1. Berichterstatter: Prof. Dr.-Ing. Alfred Mertins
2. Berichterstatter: Prof. Dr.-Ing. Erhardt Barth

Tag der mündlichen Prüfung: 29.03.2010
Zum Druck genehmigt, Lübeck, den 29.03.2010

Vorwort

Diese Arbeit entstand während meiner Tätigkeit als Wissenschaftlicher Mitarbeiter am Institut für Signalverarbeitung und Prozessrechentechnik an der Universität zu Lübeck. Erste Grundlagen wurden während meiner Tätigkeit als Wissenschaftlicher Mitarbeiter am Institut für Physik, Universität Oldenburg, gelegt.

An dieser Stelle möchte ich mich bei allen bedanken, die diese Arbeit ermöglicht haben. Das ist vor allem Prof. Dr.-Ing. A. Mertins, der mir die Möglichkeit für diese Forschungsarbeit gegeben hat. Er stand mir mit Rat und Tat während der gesamten Zeit zur Seite. Ohne diese Unterstützung wäre diese Arbeit nie entstanden.

Weiter möchte ich mich bei allen bedanken, die bei der Korrektur geholfen haben. Dies ist insbesondere Isabel Möllner, die meine Konsekutivsätze bekämpft hat, und ganz viele ‚insbesondere‘ gestrichen hat. Den letzten Schliff hat Frau Christiane Ehlers der Arbeit verpasst. Nicht zu vergessen sind die Kommentare von Alexandru Condurache, Thorsten Schmidt, Martin Panusch und Ole Jungmann. Nicht zu letzt danke ich auch Florian Müller, der im entscheidenden Moment spontan meine Lehrverpflichtung übernommen hat.

Weiter möchte ich mich bei allen Mitarbeitern und Studenten des Instituts bedanken, die meine Arbeit dort sehr interessant, abwechslungsreich aber manchmal auch anstrengend gemacht haben. Es lebe der *flat friday*.

Ein großer Dank gilt der TNF der Universität zu Lübeck, sowie der DFG für die nötige finanzielle Unterstützung.

Dankeschön!

Lübeck, im Februar 2010

Radoslaw Mazur

Inhaltsverzeichnis

1	Einleitung	1
1.1	Übersicht	1
1.2	Aufbau der Arbeit	3
1.3	Notation	4
2	Blinde Quellentrennung	7
2.1	Instantane Systeme	7
2.1.1	Die Quellen und das Mischungssystem	7
2.1.2	Das Entmischungssystem	9
2.2	Konvolutive Mischungen	15
2.3	Einsatzmöglichkeiten der blinden Quellentrennung	16
2.3.1	Akustische Mischungen	16
2.3.2	Veränderliche Mischungssysteme	17
2.3.3	Medizinische Anwendungen	18
2.4	Zusammenfassung	19
3	Independent Component Analysis	21
3.1	Unkorreliertheit und Unabhängigkeit	22
3.1.1	PCA als Vorverarbeitung	23
3.1.2	Bestimmung der Anzahl der Quellen mittels PCA	27
3.1.3	Whitening	28
3.2	ICA basierend auf der Maximierung der Kurtosis	29
3.2.1	Kurtosis als Maß für die Gaußheit	34
3.2.2	Gradientenverfahren	35
3.2.3	Fixpunkt Algorithmus	38
3.2.4	Alternativen zur Messung der Gaußheit	40
3.3	ICA basierend auf der Kullback-Leibler-Divergenz	41
3.3.1	Definitionen	42
3.3.2	Gradientenverfahren	45
3.3.3	Aktivierungsfunktionen	49
3.3.4	Komplexwertige Signale	58
3.4	Zusammenfassung	61

4	Konvolutive Mischungen	63
4.1	Die Raumimpulsantwort	64
4.2	Die Kurzzeit-Fouriertransformation	68
4.3	Konvolutives Misch- und Entmischungsmodell	73
4.3.1	Existenz des inversen Systems	74
4.3.2	Das MIMO-Entmischungssystem	77
4.4	Bewertungsmethoden	82
4.5	Zusammenfassung	84
5	Das Skalierungsproblem	87
5.1	Einfache Lösungen	89
5.2	Inverse Nachfilter	91
5.3	Max-SIR Methode	96
5.3.1	Blinde Bestimmung der Trennleistung	97
5.3.2	Berechnung der Skalierungskoeffizienten	104
5.4	Filterverkürzung	108
5.4.1	Entmischungsfiler im Zeitbereich	109
5.4.2	Filter-Optimierung	111
5.4.3	Beispiel	113
5.5	Filterformung	116
5.5.1	Optimierung der Entmischungsfiler	117
5.5.2	Fensterfunktionen	118
5.5.3	Beispiel	122
5.6	Zusammenfassung	125
6	Das Permutationsproblem	127
6.1	Nicht-blinde Verfahren	129
6.1.1	Entpermutation nach der Trennleistung	129
6.1.2	Entpermutation durch Korrelation mit den Eingangssignalen	130
6.2	Entpermutationsalgorithmen auf Basis getrennter Signale	132
6.2.1	Entpermutation auf Basis der Korrelation	132
6.2.2	Dyadisches Sortieren	135
6.2.3	Zusammenfassung	138
6.3	$\alpha\beta$ -Algorithmus	139
6.3.1	Clusterberechnung	141
6.4	Zusammenfassung	150
7	Zusammenfassung und Ausblick	151

Einleitung

1.1 Übersicht

Blinde Quellentrennung (Blind Source Separation, BSS) ist eine Technik, mit deren Hilfe aus einer Mischung von Signalen die Originalquellen bestimmt werden können. Das Verfahren heißt blind, da weder die Signale noch das Mischungssystem bekannt sind. Dieses Verfahren ist immer dann notwendig, wenn eine direkte Messung eines Signals nicht möglich oder technisch zu aufwändig ist. Ein typisches Problem im akustischen Bereich ist das sogenannte *Cocktail-Party-Problem* aus Abbildung 1.1. In diesem Fall existieren mehrere Quellen, wie zum Beispiel ein Sprecher und ein Radio, deren Signale dann mit Hilfe von Mikrofonen aufgezeichnet werden. Häufig ist in solchen Fällen eine direkte separate Aufnahme der einzelnen Quellen nicht möglich. Durch beliebig in einem Raum platzierte Mikrofone werden stets Überlagerungen dieser Signale aufgezeichnet. Die Schwierigkeit besteht darin, aus diesen Mischungen die einzelnen Quellen zu extrahieren, ohne deren Positionen zu kennen.

Das menschliche Gehör ist sehr gut in der Lage, auch unter vielen Quellen, einzelne interessante Signale aus solchen Mischungen zu extrahieren. Bei beeinträchtigtem Gehör geht diese Fähigkeit mitunter verloren. Daher besteht ein großes Interesse, die Trennung der Signale mittels eines Hörgeräts durchzuführen, um diese Beeinträchtigung zu mildern. Weitere Anwendungen der blinden Quellentrennung in technischen und medizinischen Bereichen werden im zweiten Kapitel noch genauer vorgestellt.

Um die Quellen zu bestimmen, muss also gleichzeitig zusätzlich entweder das Mischungssystem oder das dazu inverse Entmischungssystem bestimmt werden.

1 Einleitung

Ohne zusätzliche Informationen über das System oder die Signale ist eine eindeutige Bestimmung der Elemente des Entmischungssystems nicht durchführbar.

Eine sinnvolle Annahme an dieser Stelle ist ein lineares Mischungssystem und die statistische Unabhängigkeit der Quellen. Mit Hilfe der *Unabhängigkeitsanalyse* (Independent Component Analysis, ICA) können instantan gemischte Signale wieder getrennt werden. Bei der Verwendung der ICA kann das Entmischungssystem jedoch nur bis auf die Unsicherheiten in Reihenfolge und Skalierung bestimmt werden. Die Reihenfolge der gewonnenen Signale ist beliebig, denn eine Vertauschung der Signale untereinander hat keinen Einfluss auf deren Unabhängigkeit. Mit der gleichen Begründung können die getrennten Signale ebenfalls eine beliebige Skalierung aufweisen. Diese Unsicherheiten werden als das *äußere Permutations- und Skalierungsproblem* bezeichnet und können ebenfalls nicht ohne weitere Informationen gelöst werden.

Das Mischungssystem des angesprochenen Cocktail Party Problems ist jedoch komplizierter. Aufgrund der endlichen Laufzeit der Schallwellen und deren Reflexionen an Wänden und weiteren Gegenständen kommen die Signale verzögert und mehrfach überlagert bei den Mikrofonen an. Dieses konvolutive Mischungssystem kann nur mit Hilfe eines weiteren konvolutiven Entmischungssystems invertiert werden. Die direkte Bestimmung dieses Entmischungssystems im Zeitbereich ist extrem aufwändig, so dass ein Ansatz im Frequenzbereich deutlich erfolgversprechender ist. Hier wird die Faltung aus dem Zeitbereich zu einer Multiplikation, und die Trennung mit Hilfe der instantanen ICA kann in jedem Frequenzband unabhängig voneinander durchgeführt werden. Diese Vereinfachung hat jedoch andere Nachteile. Aufgrund der unabhängigen Trennung jedes Frequenzbandes sind diese beliebig permutiert und skaliert. Die sogenannten *inneren Permutations- und Skalierungsprobleme* müssen vor der Rücktransformation in den Zeitbereich gelöst werden.

Das Ziel dieser Arbeit ist es, neue Ansätze für die Lösung dieser beiden genannten Probleme vorzustellen. Für das Skalierungsproblem wird unter anderem eine Methode zur Maximierung der Trennleistung vorgestellt. Zwei weitere neue Methoden zielen darauf ab, die Entmischungsfiler in eine bestimmte, vorteilhafte Form zu bringen.

Für das Permutationsproblem wird ebenfalls eine neue Methode vorgestellt. Sie basiert auf der Modellierung mittels der generalisierten Gaußdichte und nutzt die Ähnlichkeit der Parameter in benachbarten Frequenzbändern.

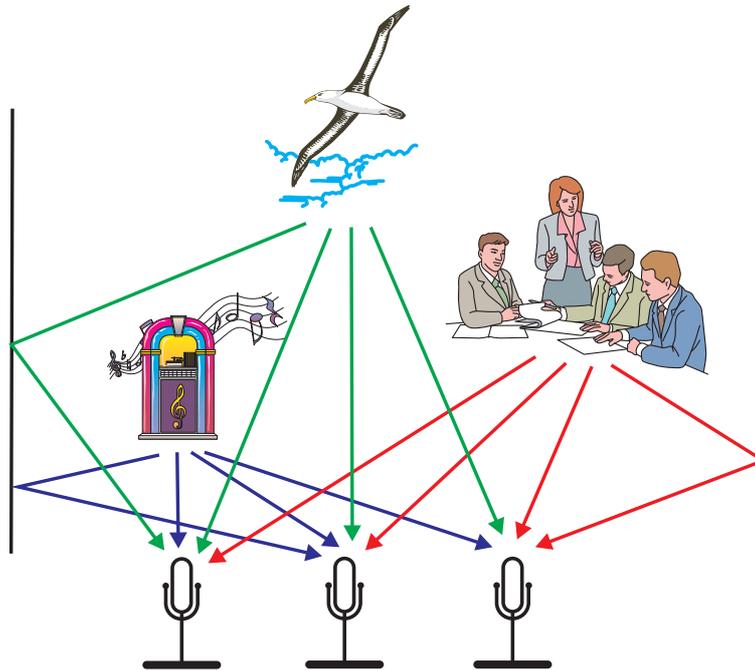


Abbildung 1.1: Das Cocktail Party Problem.

1.2 Aufbau der Arbeit

Im zweiten Kapitel dieser Arbeit wird das Misch- und Entmischungssystem vorgestellt. Dazu werden die Eigenschaften der Quellsignale sowie des Mischungssystems definiert. Daraus werden dann die notwendigen Bedingungen für ein Entmischungssystem hergeleitet. Zur Veranschaulichung werden noch einige Beispiele aus Medizin und Technik vorgestellt.

Im dritten Kapitel werden Methoden für die instantane ICA vorgestellt. Einen besonderen Schwerpunkt bilden hier Methoden, die die Nichtgaußheit der Signale ausnutzen. Zum Beispiel wird die Gaußheit mit Hilfe der Kurtosis gemessen und daraus ein Optimierungsverfahren hergeleitet. In dem zweiten Ansatz wird ein Optimierungsverfahren präsentiert, das auf einem Divergenzmaß basiert. Gemessen wird die Divergenz zwischen der gegebenen Verteilungsdichte und einer unabhängigen Verteilungsdichte, die als Referenz dient. Dieser Ansatz wird dann für komplexwertige Signale erweitert, um eine Nutzung im Frequenzbereich zu ermöglichen.

Im vierten Kapitel wird gezeigt, wie der konvolutive Fall mit Hilfe der Kurzzeit-Fourier-Transformation in den Frequenzbereich transformiert werden kann. Die Trennung wird dann instantan in jedem Frequenzband unabhängig voneinander

1 Einleitung

durchgeführt. Weiterhin werden im vierten Kapitel die Effekte des Permutations- und Skalierungsproblems genauer besprochen.

Im fünften Kapitel werden drei neue Ansätze für die Lösung des Skalierungsproblems vorgestellt. Die erste Lösung basiert auf einer blinden Schätzung der Trennleistung. Aus dieser Schätzung werden dann Skalierungskoeffizienten berechnet, so dass die Gesamttrennleistung maximiert wird. Die zweite Lösung basiert auf der Idee, die Entmischungsfiler so zu formen, dass diese einem Dirac-Impuls nachempfunden sind. Die letzte Idee erweitert diesen Ansatz, indem die Entmischungsfiler exponentiell abfallend geformt werden. Diese Form ist den Raumimpulsantworten in akustischen Umgebungen nachempfunden.

Im sechsten Kapitel wird ein neuer Ansatz für die Lösung des Permutationsproblems vorgestellt. Die zentrale Idee ist es, alle Frequenzbänder mit Hilfe der generalisierten Gaußdichte zu modellieren. Der Entpermutationsalgorithmus basiert dann auf kleinen Unterschieden der Parameter zwischen benachbarten Frequenzbändern. Eine Zusammenfassung schließt die Arbeit ab.

1.3 Notation

Innerhalb der Arbeit wurde eine konsistente Notation verwendet.

- Für Signale $x(n)$ und Filter $h(n)$ im Zeitbereich werden kleine kursive Buchstaben verwendet. Der diskrete Zeitindex ist üblicherweise n .
- Vektoren und Zufallsvektoren \mathbf{x} werden mit kleinen, fetten und kursiven Buchstaben bezeichnet. Ein mehrdimensionales Signal ist analog dazu mit $\mathbf{x}(n)$ bezeichnet.
- Für Matrizen werden große, fette Buchstaben benutzt, zum Beispiel \mathbf{A} . Ein System mit mehreren Filtern $h_{ij}(n)$ kann ebenfalls mit einer Folge von Matrizen $\mathbf{H}(n)$ beschrieben werden.
- DFT-Koeffizienten $H(\omega_k)$ eines Filters $h(n)$ werden mit Großbuchstaben gekennzeichnet. ω_k ist die diskrete Frequenz und k der Frequenzindex.
- Werden die DFT-Koeffizienten zu einer Matrix zusammengefasst, so werden fette, gerade Großbuchstaben benutzt: $\mathbf{H}(\omega_k)$. Einzelne Vektoren von $\mathbf{H}(\omega_k)$ werden dann wieder klein geschrieben: $\mathbf{h}_i(\omega_k)$.

- Eine Matrix aus $N \times M$ Nullen wird mit $\mathbf{0}_{N,M}$ abgekürzt. Bei $M = 1$ wird der zweite Index weggelassen, und es ergibt sich ein Spaltenvektor $\mathbf{0}_N$ mit N Nullen. Die Identitätsmatrix der Größe M wird mit \mathbf{I}_M abgekürzt.
- Bei Konkatenationen werden die Matrizen entsprechend untereinander geschrieben. So ist

$$\mathbf{V} = \begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}_{M,N} \end{bmatrix} \quad (1.1)$$

eine Matrix bestehend aus einer $N \times N$ Identitätsmatrix erweitert um M Zeilen mit Nullen.

- Mit \mathbf{A}^T und \mathbf{A}^H werden die Transponierte beziehungsweise die hermitesch Transponierte der Matrix \mathbf{A} bezeichnet. Eine analoge Bezeichnung wird für Vektoren benutzt. Die Inverse wird mit \mathbf{A}^{-1} und die Pseudoinverse mit \mathbf{A}^+ gekennzeichnet.
- Zufallssignale sind im Sinne von Zufallsprozessen und nicht als Musterfunktionen zu verstehen. An vielen Stellen wird statt eines Zufallssignals $\mathbf{x}(n)$ der Zufallsvektor \mathbf{x} benutzt.

1 *Einleitung*

Blinde Quellentrennung

Die Blinde Quellentrennung dient dazu, aus unbekanntem Mischungen die Originalsignale zu rekonstruieren. Blind bedeutet, dass weder das Mischsystem noch die Originalsignale bekannt sind. Ohne weitere Annahmen ist dieses Problem im Allgemeinen nicht lösbar. Eine mögliche Annahme, die zur Trennbarkeit führt, ist die statistische Unabhängigkeit der gesuchten Quellen. Bei dem im ersten Kapitel vorgestellten Cocktail-Party-Problem ist dies zum Beispiel durchaus berechtigt, da mehrere gleichzeitig sprechende Personen üblicherweise verschiedene Wörter aussprechen. Dies gilt jedoch nicht für zum Beispiel in einem Chor singende Personen, so dass hier andere Methoden angewandt werden müssen. Mit der Annahme der Unabhängigkeit kann das Verfahren der *Unabhängigkeitsanalyse* (Independent Component Analysis, ICA) angewandt werden. Die dazu nötigen Voraussetzungen werden in diesem Kapitel näher erläutert. Abschließend werden noch einige weitere Beispiele aus Medizin und Technik vorgestellt.

2.1 Instantane Systeme

2.1.1 Die Quellen und das Mischungssystem

Die einzelnen Quellen $s_i(n)$, $i = 1, 2, \dots, N$, können zu einem Vektor

$$\mathbf{s}(n) = \begin{bmatrix} s_1(n) \\ s_2(n) \\ \dots \\ s_M(n) \end{bmatrix} \quad (2.1)$$

2 Blinde Quellentrennung

der Länge M zusammengefasst werden. Neben den gesuchten Signalen können auch weitere Rauschquellen $\nu_i(n)$ existieren. Diese unbekanntes Quellen und Rauschsignale werden in einem unbekanntes System gemischt. Dieses System kann sehr unterschiedliche Eigenschaften haben, jedoch ist bei vielen physikalischen Systemen eine Annahme der Linearität möglich. Dies gilt insbesondere für das einleitend dargestellte Cocktail-Party-Problem, bei dem die Mischung durch eine Überlagerung der Schallwellen erreicht wird. In dem einfachsten Fall sind die aufgenommenen gemischten Signale Linearkombinationen der Quellen:

$$\begin{aligned}x_1(n) &= a_{11}s_1(n) + a_{12}s_2(n) + a_{13}s_3(n) \\x_2(n) &= a_{21}s_1(n) + a_{22}s_2(n) + a_{23}s_3(n) \\x_3(n) &= a_{31}s_1(n) + a_{32}s_2(n) + a_{33}s_3(n).\end{aligned}\tag{2.2}$$

Werden die gemischten Signale $x_i(n)$, $i = 1, 2, \dots, M$, zu einem Vektor

$$\mathbf{x}(n) = \begin{bmatrix} x_1(n) \\ x_2(n) \\ \dots \\ x_N(n) \end{bmatrix}\tag{2.3}$$

der Länge N und die Mischungskoeffizienten a_{ij} zu einer Matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1M} \\ a_{21} & a_{22} & \dots & a_{2M} \\ \dots & \dots & \dots & \dots \\ a_{N1} & a_{N2} & \dots & a_{N,M} \end{pmatrix}\tag{2.4}$$

der Größe $N \times M$ zusammengefasst, so kann dieses System in Matrixschreibweise als

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n)\tag{2.5}$$

zusammengefasst werden.

Das Rauschen kann auf zwei verschiedene Weisen modelliert werden. Einerseits können Rauschquellen als ganz normale Signalquellen aufgefasst werden und werden beim Trennungsprozess wieder rekonstruiert. Dieser Ansatz bedarf keiner besonderen Formulierung. Andererseits kann das Rauschen als unkorreliertes Sensorrauschen vorkommen, so dass das Modell aus Gleichung (2.5) zu

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \boldsymbol{\nu}(n)\tag{2.6}$$

mit

$$\boldsymbol{\nu}(n) = \begin{bmatrix} \nu_1(n) \\ \nu_2(n) \\ \dots \\ \nu_N(n) \end{bmatrix}, \quad (2.7)$$

dem Vektor, der das Sensorrauschen beinhaltet, erweitert werden muss.

Beispiel

In Abbildung 2.1 ist ein Beispiel für linear und instantan gemischte Signale gegeben. In Abbildung 2.1 (a) sind drei Sprachsignale abgebildet. Die Amplitude dieser Signale variiert stark, und auch Sprechpausen sind deutlich erkennbar. Alle drei Signale unterscheiden sich deutlich voneinander.

In Abbildung 2.1 (b) wurden die Signale linear nach Gleichung (2.5) gemischt. Die gemischten Signale sind einander deutlich ähnlicher und können nicht mehr so einfach unterschieden werden. Die Sprechpausen sind kaum noch erkennbar, und die Amplituden variieren nicht mehr so stark.

2.1.2 Das Entmischungssystem

In Abbildung 2.2 ist das Mischungs- und Entmischungssystem dargestellt. Auf der linken Seite sind alle Größen unbekannt, lediglich die Aufnahmen $\boldsymbol{x}(n)$ der gemischten Signale sind verfügbar. Die Aufgabe der blinden Quellentrennung ist es, ein Entmischungssystem zu entwerfen, welches die Quellen schätzt. Die Schätzung kann auf zwei verschiedene Weisen erreicht werden. Eine Möglichkeit ist es, das unbekannte Mischungssystem basierend auf $\boldsymbol{x}(n)$ zu schätzen und durch die Invertierung die originalen Quellen zu gewinnen. Da die Invertierung nicht immer möglich ist, kann stattdessen auch das Entmischungssystem direkt geschätzt werden. Dies kann zum Beispiel mit Hilfe der ICA passieren. Die Methoden hierzu werden im dritten Kapitel noch genauer besprochen.

In dem Fall einer Linearkombination der Signale aus dem letzten Abschnitt ist das Entmischungssystem

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1N} \\ b_{21} & b_{22} & \dots & b_{2N} \\ \dots & \dots & \dots & \dots \\ b_{M1} & b_{M2} & \dots & b_{M,N} \end{pmatrix} \quad (2.8)$$

2 Blinde Quellentrennung

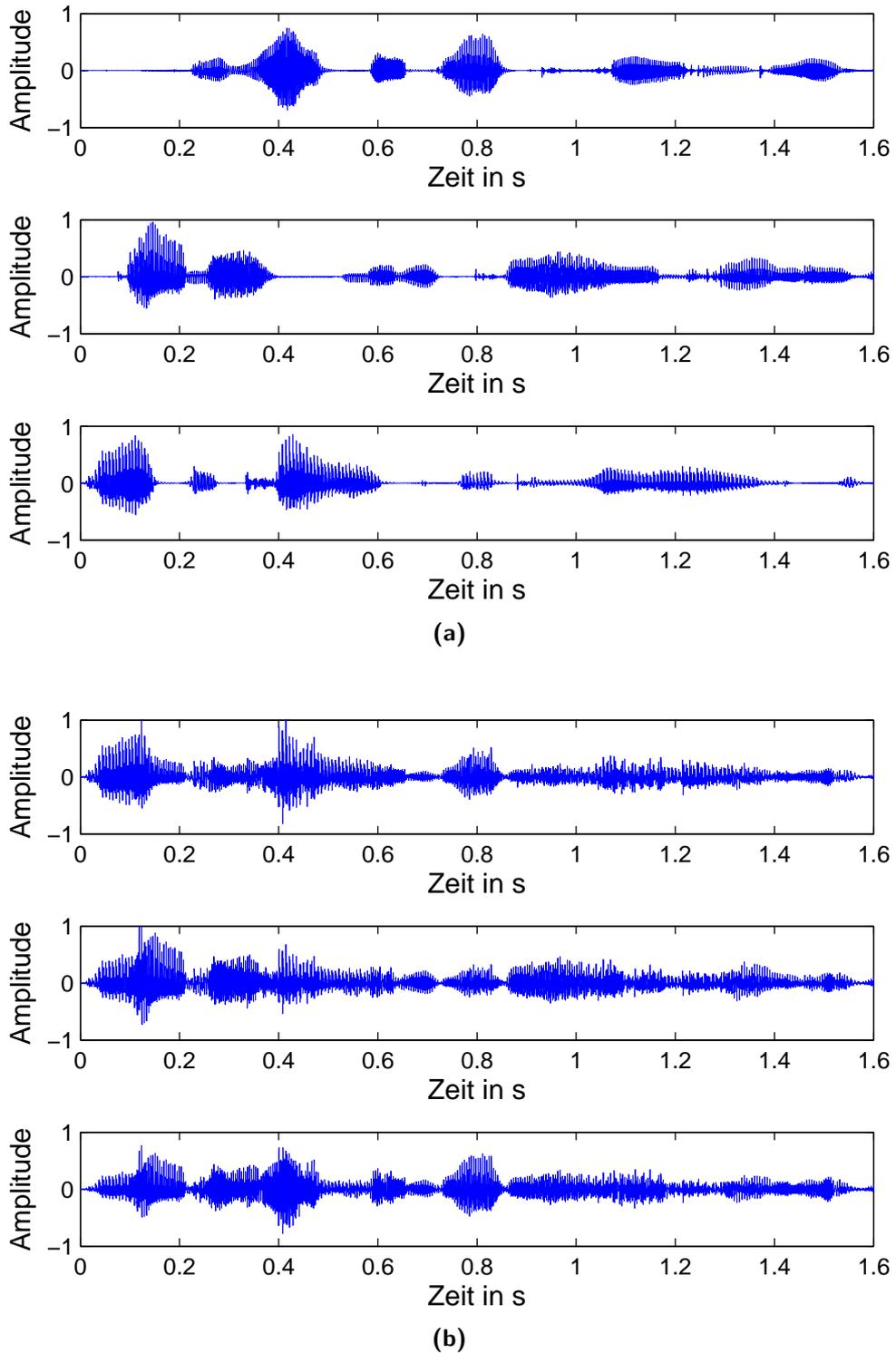


Abbildung 2.1: Beispiel für eine Mischung von Signalen. (a) Drei Sprachsignale und (b) eine lineare Mischung.

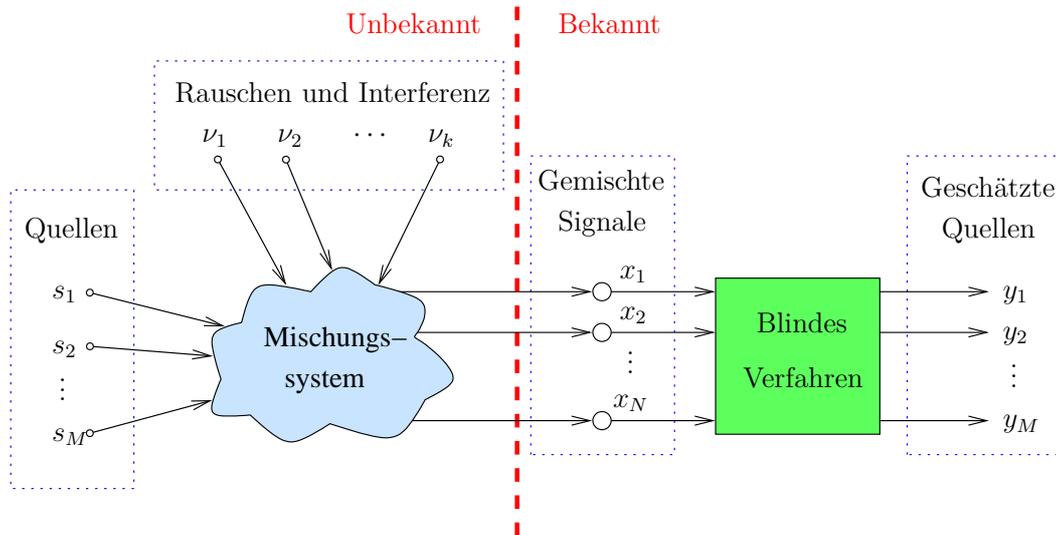


Abbildung 2.2: Beispiel für ein allgemeines Misch- und Entmischungssystem. Auf der linken Seite sind die unbekannt GröÙen dargestellt. Auf der rechten Seite befinden sich die bekannten, beziehungsweise die zu bestimmenden, Elemente.

wiederum eine Matrix, welche angewandt auf die gemischten Signale diese wieder mit

$$\mathbf{y}(n) = \mathbf{B}\mathbf{x}(n) \quad (2.9)$$

trennt. Hierbei ist

$$\mathbf{y}(n) = \begin{bmatrix} y_1(n) \\ y_2(n) \\ \dots \\ y_M(n) \end{bmatrix} \quad (2.10)$$

der Vektor, der bei erfolgreicher Trennung die geschätzten Quellen enthält.

Die Bestimmung der Entmischungsmatrix \mathbf{B} basiert auf der Annahme der Unabhängigkeit der einzelnen Signale. Diese Annahme ist jedoch nicht ausreichend, um \mathbf{B} eindeutig zu bestimmen. So hat zum Beispiel eine Skalierung oder ein Vertauschen der Signale keinen Einfluß auf deren Unabhängigkeit. Somit werden lediglich skalierte und permutierte Versionen

$$\mathbf{y}(n) = \mathbf{D}\mathbf{\Pi}\mathbf{s}(n) \quad (2.11)$$

der Quellen rekonstruiert. Hierbei ist \mathbf{D} eine Diagonal-Matrix und $\mathbf{\Pi}$ eine Permutationsmatrix. Ohne weitere Informationen über das Entmischungssystem ist eine Korrektur dieser Unsicherheiten unmöglich.

2 Blinde Quellentrennung

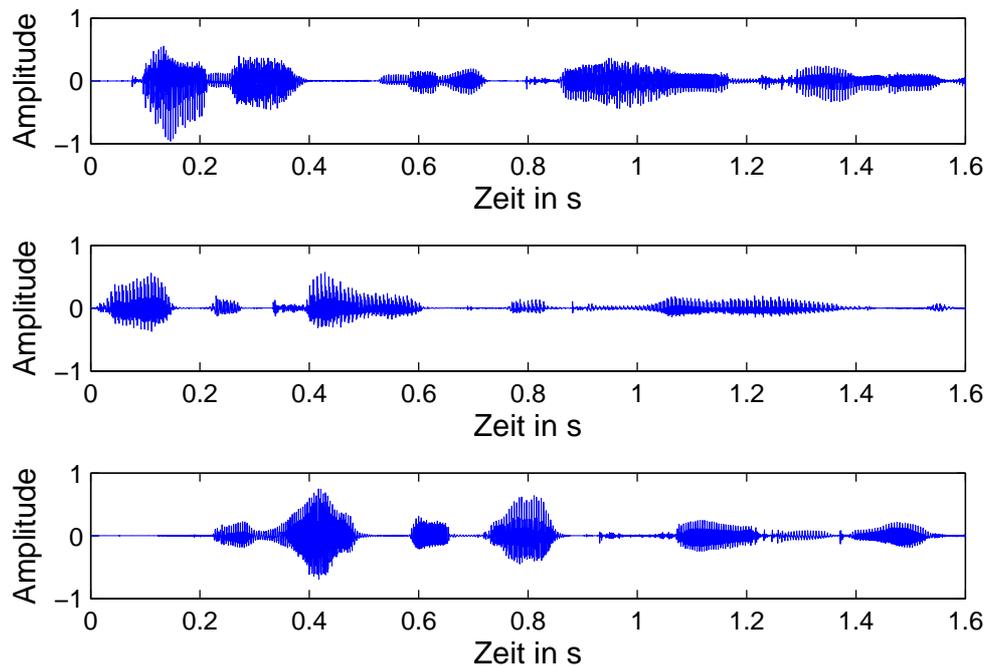


Abbildung 2.3: Die drei gefundenen Komponenten der Mischung aus der Abbildung 2.1 (b). Die Reihenfolge der Signale ist vertauscht und die Energie ist verändert.

Beispiele

Die gemischten Signale aus dem Beispiel aus Abbildung 2.1 wurden mit Hilfe einer ICA-Methode des dritten Kapitels getrennt. Die Ergebnisse sind in Abbildung 2.3 dargestellt. Die Trennung war in diesem Fall erfolgreich, aber die Uneindeutigkeit der ICA-Methode ist deutlich sichtbar. So ist die Reihenfolge der Signale vertauscht, und auch die Energie ist nicht erhalten geblieben. Die erste getrennte Quelle wurde sogar mit einem negativen Faktor multipliziert.

Die Gründe für die Unsicherheiten der ICA können anhand eines zweiten Beispiels in Abbildung 2.4 anschaulich erklärt werden. Hier ist in Abbildung 2.4 (a) das Streubild zweier gemischter Signale dargestellt. In diesem Fall sind die beiden gesuchten Komponenten deutlich an den Häufungen entlang gerader Linien sichtbar. Das Ergebnis nach der Trennung ist in Abbildung 2.4 (b) dargestellt. Hier sind die Komponenten entlang der Achsen ausgerichtet, was gleichzeitig die Unabhängigkeit dieser Komponenten bedeutet. Jede Komponente kann aber für sich einzeln skaliert werden, wie in Abbildung 2.4 (c) dargestellt, ohne dass die Unabhängigkeit verloren geht. Auch eine Vertauschung der Komponenten, wie

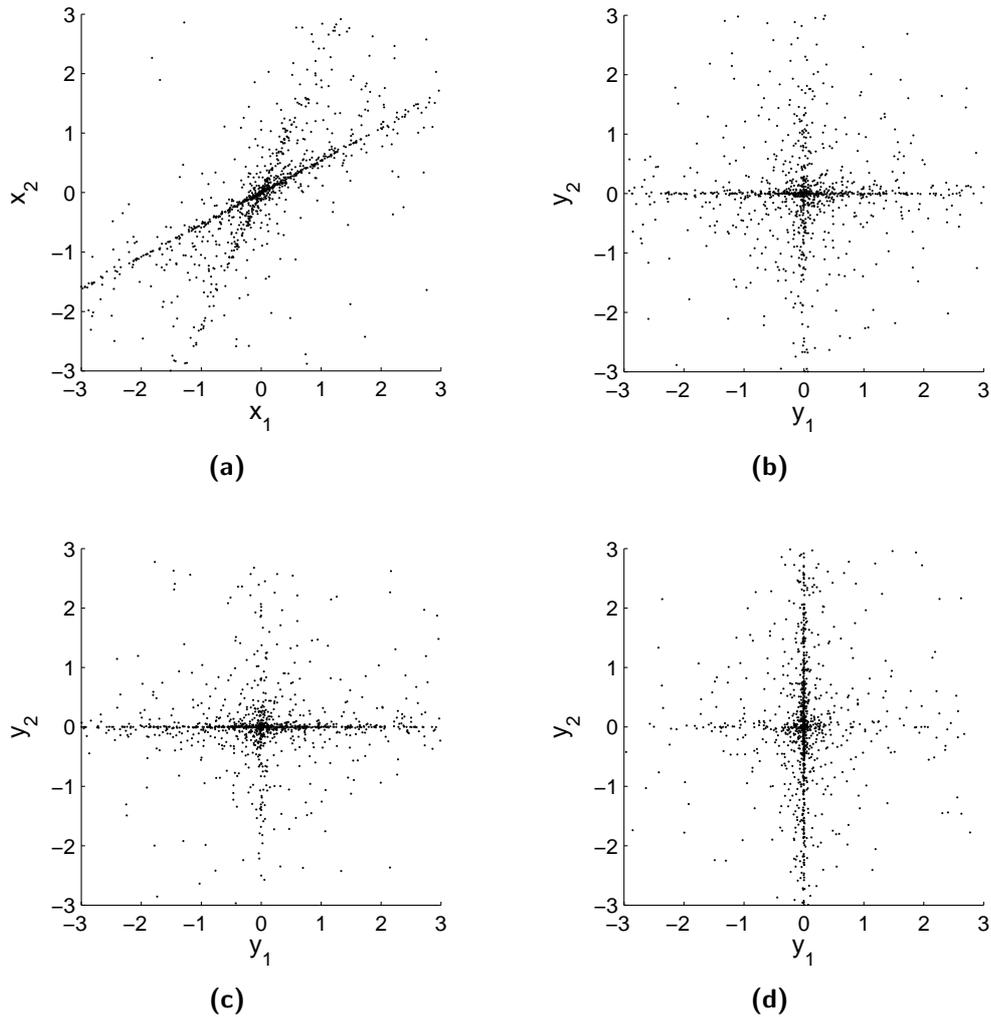


Abbildung 2.4: Die Unsicherheiten der ICA. (a) Scatterplot zweier gemischten Signale. (b) Die unabhängigen Komponenten und (c) deren Skalierung sowie (d) eine Permutation.

in Abbildung 2.4 (d) dargestellt, ändert nichts an ihrer Unabhängigkeit.

Weitere Bedingungen

Damit die Quellensignale nach Gleichung (2.9) bestimmt werden können, müssen \mathbf{A} und \mathbf{B} , welche das Misch- und Entmischungssystem repräsentieren, noch weiteren Bedingungen genügen. Zusammen mit den Skalierungen und Permu-

2 Blinde Quellentrennung

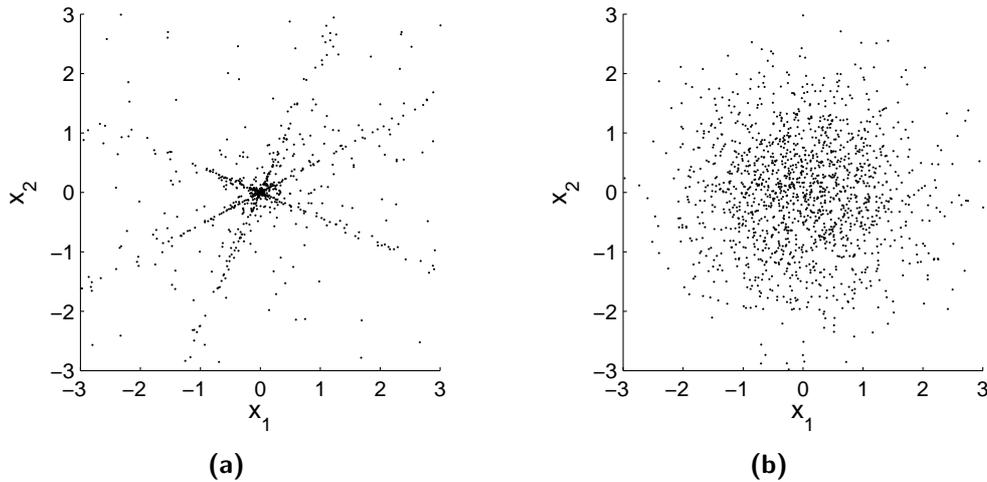


Abbildung 2.5: Gemischte Signale, die nicht mit Hilfe linearer Methoden separiert werden können. (a) Eine Mischung mit mehr Komponenten als Beobachtungen und (b) eine Mischung von Signalen mit Gaußverteilung.

tationen lautet das Gesamtsystem

$$\mathbf{BA} = \mathbf{DI}. \quad (2.12)$$

Um die Gleichung (2.12) zu erfüllen, muss \mathbf{A} mindestens so viele Zeilen wie Spalten und den maximalen Rang haben. Dies bedeutet, dass mindestens so viele Beobachtungen wie Quellen vorhanden sein müssen. Für den *kritischen Fall* der gleichen Anzahl an Beobachtungen wie Quellen, also einer quadratischen Mischungsmatrix \mathbf{A} , darf diese nicht singular sein. Ein Beispiel, in dem diese Bedingung verletzt ist, ist in Abbildung 2.5 (a) gegeben. Hier liegen zwei Beobachtungen von insgesamt drei gemischten Signalen vor. Auch wenn die einzelnen Komponenten erkennbar sind, ist eine Trennung nur mit Hilfe nicht-linearer Methoden möglich.

Ein weiterer problematischer Fall ist in Abbildung 2.5 (b) dargestellt. Hier wurden zwei Signale mit Gaußverteilung gemischt. In diesem Fall ist keine Struktur erkennbar, mit deren Hilfe eine Bestimmung der Komponenten möglich wäre.

Grundsätzlich können mit Hilfe der ICA nur Quellen bestimmt werden, die keine Gaußverteilung haben. Dies ist kein Problem, sofern nur eine einzige Quelle eine Gaußverteilung hat, denn in diesem Fall können immer noch alle nicht gaußschen Quellen bestimmt werden, so dass der Anteil mit Gaußverteilung eben dieser bestimmten Quelle zugeordnet werden kann. Bei mehr als einer

Quelle mit Gaußverteilung kann dieser Anteil nicht getrennt werden. Für die Trennung von Sprachsignalen ist dies keine große Einschränkung, da diese im Allgemeinen eher eine Laplace-Verteilung aufweisen.

2.2 Konvolutive Mischungen

Die einfachen Misch- und Entmischungsmodelle aus den Gleichungen (2.5) und (2.9) sind für den echten Fall einer akustischen Mischung nicht zutreffend. Aufgrund der endlichen Schallgeschwindigkeit und möglichen Reflexionen an Wänden und weiteren Gegenständen treffen die Signale mehrfach und unterschiedlich verzögert auf die Mikrofone auf. Dies kann mit Hilfe der Raumimpulsantworten $h_{ij}(n)$ modelliert werden, die in realistischen Szenarios mehrere tausend Koeffizienten lang sind. Die Aufnahmen $x_j(n)$, $j = 1, 2, \dots, N$, sind somit Überlagerungen von gefilterten Quellen:

$$x_j(n) = \sum_{i=0}^{N-1} h_{ij}(n) * s_i(n). \quad (2.13)$$

Mit

$$\mathbf{H}(n) = \begin{pmatrix} h_{11}(n) & h_{12}(n) & \dots & h_{1M}(n) \\ h_{21}(n) & h_{22}(n) & \dots & h_{2M}(n) \\ \dots & \dots & \dots & \dots \\ h_{N1}(n) & h_{N2}(n) & \dots & h_{N,M}(n) \end{pmatrix} \quad (2.14)$$

lautet das konvolutive Mischungsmodell

$$\mathbf{x}(n) = \mathbf{H}(n) * \mathbf{s}(n). \quad (2.15)$$

Im vierten Kapitel wird gezeigt werden, wie dieses Mischungssystem wiederum mit Hilfe von Entmischungsfilttern $w_{ij}(n)$ mit

$$\mathbf{W}(n) = \begin{pmatrix} w_{11}(n) & w_{12}(n) & \dots & w_{1M}(n) \\ w_{21}(n) & w_{22}(n) & \dots & w_{2M}(n) \\ \dots & \dots & \dots & \dots \\ w_{N1}(n) & w_{N2}(n) & \dots & w_{N,M}(n) \end{pmatrix} \quad (2.16)$$

invertiert werden kann. Das konvolutive Entmischungssystem lautet damit:

$$\mathbf{y}(n) = \mathbf{W}(n) * \mathbf{x}(n). \quad (2.17)$$

Da eine direkte Bestimmung der Entmischungsfiltter $w_{ij}(n)$ sehr aufwändig ist, kann eine Trennung im Frequenzbereich durchgeführt werden. So wird aus der

2 *Blinde Quellentrennung*

Faltung im Zeitbereich eine Multiplikation im Frequenzbereich, bei der jede Frequenz mit Hilfe der instantanen ICA getrennt werden kann. Der Nachteil dieser Methode ist die zufällige Reihenfolge und Skalierung der getrennten Signale in jeder Frequenz. Wird die Skalierung vor der Rücktransformation nicht korrigiert, so werden lediglich, unter Umständen stark, gefilterte Versionen rekonstruiert. Eine fehlende Korrektur der Permutationen ist noch problematischer, da in diesem Fall mehrere Signale in verschiedenen Frequenzbändern einem Ausgang zugeordnet werden und die Trennung somit fehlschlägt.

Die Transformation in den Frequenzbereich wird im vierten Kapitel noch genauer vorgestellt. Neue Ansätze für die Lösung des Skalierungsproblems werden im fünften und für das Permutationsproblem im sechsten Kapitel besprochen.

2.3 Einsatzmöglichkeiten der blinden Quellentrennung

2.3.1 Akustische Mischungen

Das Cocktail-Party-Problem, bei dem mehrere Personen gleichzeitig sprechen, wurde bereits angesprochen. Die Lösung dieses Problems ist in diversen Situationen von Interesse. Das beste Beispiel ist hier die Vorverarbeitung in einem Hörgerät. Personen mit Defiziten der Hörleistung können oftmals erhebliche Probleme haben, Gesprächen zu folgen, wenn viele Nebengeräusche existieren. So könnte mit Hilfe der blinden Quellentrennung die relevante Quelle extrahiert werden.

Für eine Implementierung müssen aber noch weitere Probleme gelöst werden. Bereits die Frage, welche Stimme die relevante ist, ist nicht trivial zu beantworten. Ein möglicher Ansatz hier ist ein halb-blindes Verfahren: Bei bekannter Lage der Mikrofone könnte zum Beispiel die Stimme ausgewählt werden, die von vorne kommt, da diese mit großer Wahrscheinlichkeit von dem gewünschten Gesprächspartner stammt.

Weitere Probleme ergeben sich aus der Tatsache, dass das Mischungssystem sich ständig ändert. So reichen bereits kleine Bewegungen der Personen oder auch nur ein leichtes Verdrehen des Kopfes, um die Impulsantworten so stark zu verändern, dass eine Anpassung notwendig wird. Diese kann zwar durch adaptive Verfahren realisiert werden, aber die Adaptationsgeschwindigkeit ist oftmals zu gering.

2.3 Einsatzmöglichkeiten der blinden Quellentrennung

Eine weitere akustische Anwendung sind Freisprechanlagen. Ausgehend von einer Benutzung in einem PKW ergeben sich hier andere Probleme. Zwar ist die Bewegungsfreiheit der Insassen eines PKWs deutlich eingeschränkt, so dass die Richtungsinformationen besser ausgewertet werden können, aber die Nebengeräusche sind oftmals hoch und können keiner besonderen Richtung zugeordnet werden.

Die akustische Überwachung von mechanischen Maschinen ist ein weiterer möglicher Einsatz für die blinde Quellentrennung. So produzieren verschiedene Teile einer Maschine unterschiedliche Geräusche. Ändert sich eines davon, so deutet das auf ein mögliches technisches Problem hin. Mit Hilfe der blinden Quellentrennung können diese Geräusche einzeln isoliert und überwacht werden.

2.3.2 Veränderliche Mischungssysteme

Eine Vorverarbeitung der Signale, die in sich verändernden Systemen gemischt wurden, kann ebenfalls mit Hilfe der blinden Quellentrennung durchgeführt werden. Ein Beispiel hierfür sind Straßensensoren, die sowohl Verkehrsdichte als auch die Art der Fahrzeuge bestimmen sollen. Unter der Fahrbahn installierte Sensoren messen hierbei die Änderung des Magnetfeldes, welche durch die darüber fahrenden Fahrzeuge verursacht wird. Je nach Art und Position der Sensoren kann ein Fahrzeug von mehreren Sensoren erfasst werden. Bei mehreren Spuren kann ein Sensor andererseits überlagerte Signale aufnehmen, die von mehreren Fahrzeugen verursacht wurden.

Für eine einfache Verkehrszählung sind diese Signale oftmals ausreichend. Für die Bestimmung der Fahrzeugart mit Hilfe eines Klassifikators müssen die Signale jedoch getrennt vorliegen. Die besondere Schwierigkeit dieser Trennung liegt in dem zeitlich veränderlichen Mischungsmodell

$$x_i(t) = \sum_{j=1}^N \int_{-\infty}^{\infty} s_j(t \cdot v_j(t) - \tau) \cdot h_{ij}(\tau) d\tau \quad (2.18)$$

mit v_j der Geschwindigkeit der Fahrzeuge, $h_{ij}(p)$ den örtlichen Sensitivitätsprofilen der Sensoren und einer örtlichen Fahrzeugsignatur $s_j(p)$, welche geschwindigkeitsabhängig in zeitliche Profile transformiert werden. Eine blinde lineare Entmischung ist nicht möglich. Mit zusätzlichen Informationen, wie zum Beispiel der Geschwindigkeit der Fahrzeuge, kann ein zeitlich variierendes Entmischungssystem entworfen werden.

2.3.3 Medizinische Anwendungen

Geräusche der inneren Organe

Die Signale des menschlichen Körpers werden in der Medizin in vielfältiger Art für die Diagnostik benutzt. So produzieren die inneren Organe, wie zum Beispiel das Herz, die Lunge oder der Verdauungstrakt Geräusche. Mit nichtinvasiven Methoden können diese Geräusche nicht einzeln aufgenommen werden. Die Mischungen, die mit Hilfe der auf der Haut angebrachten Mikrofone aufgenommen werden, können mit der blinden Quellentrennung getrennt werden. Die besondere Schwierigkeit ist hier die ständige Änderung des Mischungssystems, verursacht durch Atmung oder Bewegungen des Verdauungstraktes.

Ein besonders interessanter Fall ist bei der pränatalen Diagnostik gegeben. Die Herztöne des Fötus können nicht unabhängig von denen der Mutter gehört werden. Die Trennung dieser Signale ist besonders schwierig, da neben den oben genannten Problemen zusätzlich der große Unterschied der Lautstärken beachtet werden muss. So haben die schützenden Schichten, die den Fötus umgeben, eine sehr starke Dämpfung der akustischen Signale zur Folge.

EKG und EEG

Muskeln werden mit Hilfe von elektrischen Signalen gesteuert. Besonders interessant sind hier zum Beispiel die periodischen Signale, die das Herz zum Schlagen anregen. Diese Signale können mit Hilfe des Elektrokardiogramms (EKG) aufgenommen und ausgewertet werden.

Bei Schwangeren können diese Signale wiederum nur als Mischung der Signale der Mutter und des Kindes aufgenommen werden. Die Methoden der blinden Quellentrennung können hier eingesetzt werden, um die Trennung dieser Signale zu erreichen.

Bei der Elektroenzephalographie (EEG) wird mit Hilfe mehrerer Elektroden die Gehirnaktivität gemessen. Werden in einem Bereich des Gehirns gleichzeitig viele Neuronen aktiv, so führt dies zu einer Ladungsverschiebung, die zu einem elektrischen Feld führt. Dieses Feld kann mit Elektroden auf der Kopfoberfläche gemessen werden. Da im Gehirn üblicherweise mehrere solcher Regionen existieren, zeichnen die Elektroden wiederum eine Mischung dieser Signale auf.

Neben der Bestimmung der einzelnen Signale ist in diesem Fall auch die Position und eine mögliche Bewegung dieser aktiven Regionen interessant. Diese kann aus

den Positionen der einzelnen Elektroden berechnet werden. Das Verfahren ist in diesem Fall nicht vollständig blind.

2.4 Zusammenfassung

In diesem Kapitel wurde gezeigt, wie die blinde Quellentrennung mit Hilfe der Unabhängigkeitsanalyse gelöst werden kann. Der Ansatz, die einzelnen Quellen als unabhängig anzunehmen, ist ausreichend um die Mischungen zu trennen. Jedoch ist dieser Ansatz mit dem Nachteil der unbekanntem Reihenfolge und Skalierungen der rekonstruierten Signale verbunden.

Die akustische Mischung von Signalen kann mit Hilfe eines konvolutiven Mischungsmodells beschrieben werden. Im Frequenzbereich kann dieses auf mehrere instantane Probleme reduziert werden. Im nächsten Kapitel werden Methoden vorgestellt, mit denen diese instantanen Mischungen getrennt werden können.

2 *Blinde Quellentrennung*

Kapitel 3

Independent Component Analysis

Das statistisch-mathematische Verfahren für die instantane blinde Quellentrennung ist die Independent Component Analysis (ICA). Es existieren verschiedene Ansätze, die je nach der Methode zur Messung der Unabhängigkeit zu unterschiedlichen Algorithmen führen.

Die instantane ICA wird aus zwei Gründen betrachtet. Einerseits wird so die grundlegende Einsicht für alle blinden Trennverfahren gelegt. Andererseits kann der für diese Arbeit entscheidende konvolutive Fall auf mehrfache, instantane Probleme im Zeit-Frequenz-Bereich zurückgeführt werden. Durch diese Transformation werden die reellwertigen Eingangswerte allerdings in komplexwertige überführt, so dass auch insbesondere der komplexwertige Fall hier betrachtet werden muss.

In diesem Kapitel werden Algorithmen vorgestellt, die geeignet sind, instantane Mischungen zu trennen. Bei den hier vorgestellten Algorithmen wird die Nichtgaußheit der Signale ausgenutzt. So haben Mischungen von unabhängigen Signalen eine Verteilung, die näher an der Gaußverteilung ist als die einzelnen Signale. Durch die Messung der Gaußheit mit Hilfe der *Kurtosis* können einzelne Quellen identifiziert und mit Hilfe einer Projektion extrahiert werden.

Ein weiterer Ansatz basiert auf einem informationstheoretischen Maß. Dabei wird die Unabhängigkeit der Signale durch die Separierbarkeit ihrer Verbundverteilungsdichten ausgedrückt. Dieser Ansatz führt zu dem Bell-Sejenowski-Algorithmus [9], der sehr effizient ist und auf komplexwertige Signale erweitert werden kann. Eine weitere Verbesserung liegt in der Nutzung des natürlichen Gradienten [5, 3], der die Konvergenzgeschwindigkeit zusätzlich steigert.

3 Independent Component Analysis

Weitere Algorithmen nutzen zum Beispiel die Nichtstationarität der Signale aus. In [67] wurde gezeigt, dass durch das gemeinsame Diagonalisieren der Korrelationsmatrizen aus verschiedenen Epochen eine Trennung erreicht werden kann. Sind die Signale stationär und besitzen unterschiedliche Leistungsdichten, so kann das Diagonalisieren der Korrelationsmatrizen mit unterschiedlichen Verzögerungen zur Bestimmung der Entmischungsmatrix benutzt werden [59].

Da Sprachsignale nichtstationär sind, eine stark nichtgaußsche Statistik besitzen und zudem spärlich in der Zeit-Frequenz-Ebene repräsentiert sind, eignen sich prinzipiell verschiedene der oben genannten Algorithmen für die Trennung von Sprachsignalen. Im Rahmen dieser Arbeit werden jedoch nur die auf einer Beobachtung der Amplitudenstatistik basierenden Verfahren berücksichtigt, weil sie sich besonders vorteilhaft von instantanen auf konvolutive Mischungen erweitern lassen und bereits bei relativ kurzen beobachteten Signalen konvergieren. Bevor diese Methoden näher beschrieben werden, wird jedoch mit der Principal Component Analysis (PCA) eine Methode vorgestellt, die eine Vorverarbeitung erlaubt, die die eigentliche Trennung erheblich vereinfachen kann.

3.1 Unkorreliertheit und Unabhängigkeit

Oftmals wird die Unkorreliertheit von Zufallsvariablen gleichgesetzt mit deren Unabhängigkeit. Dies ist eine irriige Annahme, da die Unabhängigkeit eine viel stärkere Eigenschaft ist. Das kann bereits an einfachen Beispielen gezeigt werden, wie in Abbildung 3.1 dargestellt. In dem ersten Beispiel ist eine Mischung aus zwei supergaußschen und in dem zweiten aus zwei subgaußschen Signalen¹ $\mathbf{x}(n)$ als Streubild dargestellt. In Abbildung 3.1 (a) sind die Komponenten an Häufungen entlang gerader Linien und in Abbildung 3.1 (b) entlang der Ränder erkennbar. Die Komponenten sind nicht entlang der Achsen ausgerichtet, so dass die Signale $x_i(n)$ voneinander abhängig sind. Trotzdem ist in beiden Fällen die Korrelationsmatrix

$$\mathbf{R}_{xx} = E\{\mathbf{x}(n)\mathbf{x}(n)^T\} = \mathbf{I} \quad (3.1)$$

mit $E\{\cdot\}$, dem Erwartungswert, bereits diagonal. Das heißt, die Signale $x_i(n)$ sind in diesen Beispielen unkorreliert.

In Abbildung 3.2 wurden die Signale so verändert, dass die Komponenten entlang der Achsen ausgerichtet und somit unabhängig sind. Die Korrelationsmatrix

¹Definition für super- und subgaußisch siehe Abschnitt 3.2.1

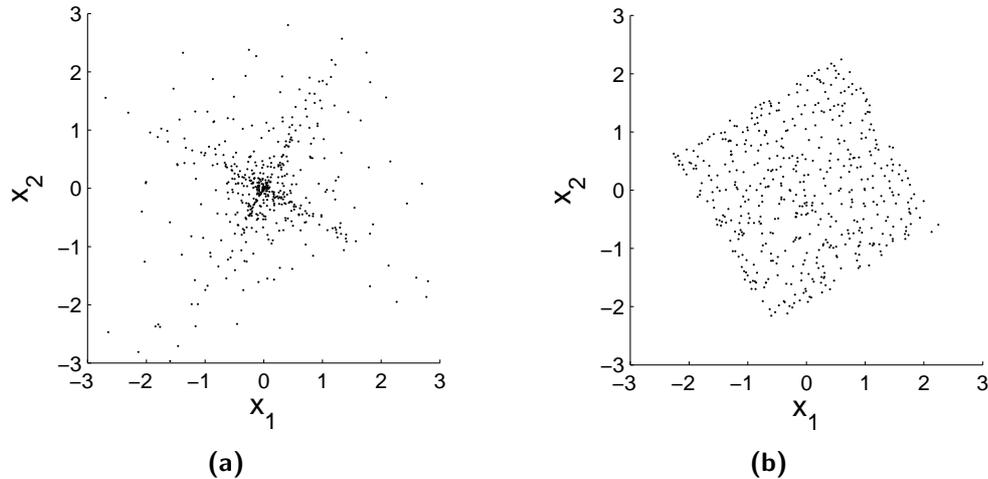


Abbildung 3.1: Zwei Beispiele für unkorrelierte Datensätze. (a) Eine Mischung aus zwei supergaußschen und (b) subgaußschen Signalen. In beiden Fällen ist die Korrelationsmatrix diagonal.

ist aber unverändert geblieben. Alleine mittels einer Statistik zweiter Ordnung kann die Trennung der Signale nicht erreicht werden. Eine Dekorrelation alleine führt somit nicht zu unabhängigen Komponenten.

3.1.1 PCA als Vorverarbeitung

Ein oft benutztes Verfahren für die Untersuchung von Daten ist die Hauptachsentransformation [68]. Dieses Verfahren ist auch bekannt als Hotelling-Transformation oder Karhunen-Loeve-Transformation [36]. Diese Transformation hat zwei grundsätzliche Ziele. Einerseits sollen Signale dekorreliert werden und andererseits, sofern möglich, in eine niedrigdimensionale Darstellung transformiert werden.

Das erste Ziel kann durch ein zweistufiges Verfahren erreicht werden. Zuerst werden die Signale durch die Subtraktion des Mittelwertes

$$\tilde{\mathbf{x}}(n) = \mathbf{x}(n) - E\{\mathbf{x}(n)\} \quad (3.2)$$

mittelwertfrei gemacht. Danach werden die Signale mit Hilfe einer orthonormalen Transformationsmatrix \mathbf{V} so gedreht, dass die einzelnen Komponenten von

$$\mathbf{y}(n) = \mathbf{V}^T \mathbf{x}(n) \quad (3.3)$$

3 Independent Component Analysis

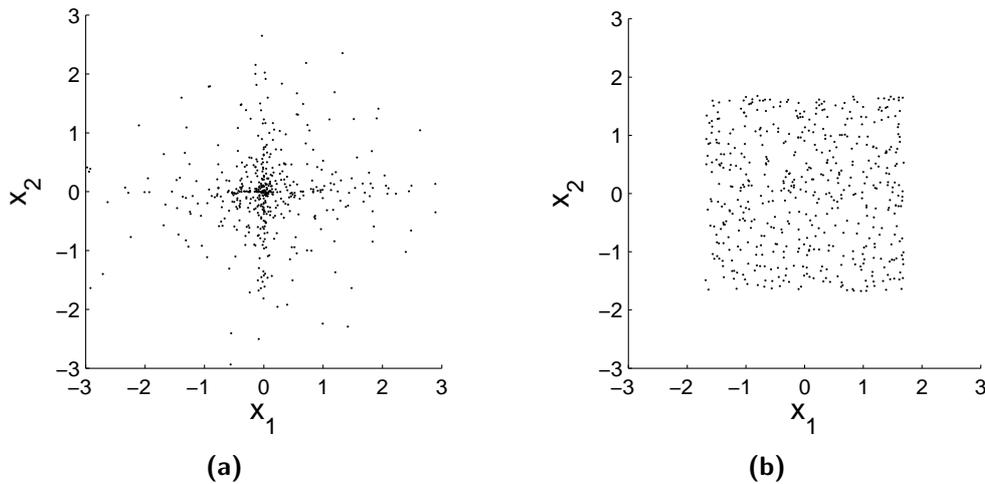


Abbildung 3.2: Zwei Beispiele für unabhängige Datensätze. Die Daten aus Abbildung 3.1 wurden so gedreht, dass sie unabhängig sind. Die Korrelationsmatrix ist dabei unverändert geblieben.

unkorreliert sind, das heißt, dass

$$E\{y_i(n) y_j(n)\} = \delta_{ij} \quad (3.4)$$

mit δ_{ij} , dem Kronecker-Delta, gilt. Ein Beispiel für die beiden Schritte dieser Transformation ist in Abbildung 3.3 gegeben.

Das zweite Ziel dieser Transformation ist es, eine niedrigdimensionale Darstellung der Signale zu erhalten. Durch das Weglassen einzelner Signale $y_i(n)$, die nur einen kleinen Beitrag zu $\mathbf{y}(n)$ leisten, kann eine reduzierte Darstellung erreicht werden. In dem Beispiel aus Abbildung 3.3 würde die erste Komponente mehr über $\mathbf{y}(n)$ aussagen als die zweite. Demzufolge wäre die erste Komponente die beste eindimensionale Beschreibung des mehrdimensionalen Signals. Dies kann dann beispielsweise für eine Kompression benutzt werden.

Diese beiden Teilverfahren haben, grob gesehen, eine ähnliche Zielsetzung wie die ICA. Allerdings ist, wie vorher gezeigt, eine Dekorrelation im Allgemeinen nicht ausreichend für die Bestimmung von unabhängigen Signalen. Das Verfahren der PCA kann jedoch als Vorverarbeitungsschritt benutzt werden, da sie die nachfolgenden Schritte der ICA vereinfachen kann.

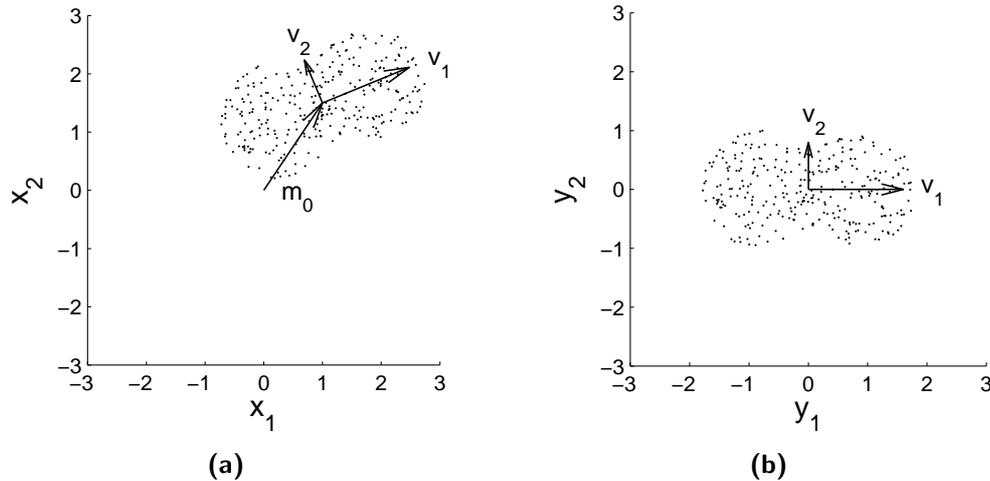


Abbildung 3.3: Principal Component Analysis (a) Die Verteilung der Daten wird durch den Mittelpunkt \mathbf{m}_0 und die beiden Hauptkomponenten \mathbf{v}_1 und \mathbf{v}_2 beschrieben. (b) Nach der Subtraktion des Mittelwertes und einer orthogonalen Transformation sind die Hauptkomponenten entlang der Koordinatenachsen ausgerichtet.

Berechnung der Dekorrelationsmatrix

Die Bestimmung der Transformationsmatrix \mathbf{V} , die zu dekorrelierten Signalen nach Gleichung (3.3) führt, kann auf ein Eigenwertproblem zurückgeführt werden. Dazu wird die Korrelationsmatrix² \mathbf{R}_{xx} mittels

$$\mathbf{R}_{xx} = E\{\mathbf{x}(n)\mathbf{x}(n)^T\} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \quad (3.5)$$

diagonalisiert. Hierbei ist

$$\mathbf{\Lambda} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_N\} \quad (3.6)$$

eine diagonale Matrix mit den Eigenwerten der Korrelationsmatrix auf der Hauptdiagonalen und \mathbf{V} eine quadratische Matrix bestehend aus den zugehörigen Eigenvektoren \mathbf{v}_i der Länge eins:

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N]. \quad (3.7)$$

Da die Eigenvektoren senkrecht zueinander stehen, wird auch die Orthogonalitätsbedingung an \mathbf{V} mit

$$\mathbf{v}_i^T \mathbf{v}_j = \delta_{ij} \quad (3.8)$$

²Bei mittelwertfreien Daten ist die Korrelationsmatrix gleich der Kovarianzmatrix.

3 Independent Component Analysis

erfüllt.

Die Bestimmung von \mathbf{V} kann auf die Lösung des Eigenwertproblems

$$\mathbf{R}_{xx}\mathbf{v}_i = \lambda_i\mathbf{v}_i \quad (3.9)$$

zurückgeführt werden. Eine Besonderheit ergibt sich aus der Symmetrie von \mathbf{R}_{xx} . Das bedeutet, dass \mathbf{R}_{xx} positiv definit ist und somit alle λ_i nicht negativ sein können.

Die übliche Vorgehensweise an dieser Stelle wäre also die Berechnung der Korrelationsmatrix und die Lösung des Eigenwertproblems. Die Lösung des Eigenwertproblems ist relativ unproblematisch, da hierfür effiziente Algorithmen existieren. Dies gilt auch für Fälle mit Daten, die mehrere Tausend Dimensionen haben und deren Korrelationsmatrix demzufolge entsprechend groß ist.

Die Berechnung der Korrelationsmatrix ist dagegen nicht immer unproblematisch. Im Allgemeinen ist die echte Korrelationsmatrix \mathbf{R}_{xx} nicht bekannt, sondern nur eine Schätzung $\widehat{\mathbf{R}}_{xx}$ basierend auf einer endlich großen Stichprobemenge:

$$\widehat{\mathbf{R}}_{xx} = \frac{1}{M} \sum_{n=1}^M \mathbf{x}(n)\mathbf{x}(n)^T. \quad (3.10)$$

Ist der Prozess, der zu den Signalen $\mathbf{x}(n)$ führt, nicht stationär, so kann sich die Korrelationsmatrix mit der Zeit ändern. In diesem Fall kann ein adaptives Verfahren für die zeitabhängige Bestimmung von $\widehat{\mathbf{R}}_{xx}$ benutzt werden. Hierzu kann die Online-Regel

$$\widehat{\mathbf{R}}_{xx}^{(n+1)} = (1 - \beta_0) \cdot \widehat{\mathbf{R}}_{xx}^{(n)} + \beta_0 \cdot \mathbf{x}(n)\mathbf{x}(n)^T \quad (3.11)$$

mit β_0 als Faktor, der die Adaptationsgeschwindigkeit regelt, benutzt werden.

Es existieren noch weitere Algorithmen für die Bestimmung der Hauptkomponenten. Eine zentrale Idee ist es, die Berechnung der Korrelationsmatrix zu umgehen und die Vektoren \mathbf{v}_i direkt zu bestimmen. [64, 62, 24]. (Eine Übersicht findet sich in [18].) Für den in dieser Arbeit hauptsächlich betrachteten Fall von einigen wenigen (akustischen) Mischungen ist die oben beschriebene Vorgehensweise aber ausreichend.

3.1.2 Bestimmung der Anzahl der Quellen mittels PCA

Die Reihenfolge der Eigenvektoren aus Gleichung (3.9) ist zufällig. Werden jedoch die Eigenwerte (und die dazugehörigen Eigenvektoren) in der Matrix $\mathbf{\Lambda}$ absteigend sortiert

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N, \quad (3.12)$$

so bekommen die einzelnen Komponenten eine weitere Bedeutung. Der Betrag von λ_i gibt nämlich an, wie gut die Komponente i die Daten beschreibt. Durch die Sortierung werden somit die Komponenten, die besonders gut die Originaldaten repräsentieren, nach vorne geschoben. Bei der Karhunen-Loeve-Transformation wird dieser Effekt ausgenutzt, um mit Hilfe eines Teils der Komponenten die gesamten Daten möglichst gut anzunähern. Werden die ersten M Eigenvektoren, die zu den größten Eigenwerten λ_i gehören, in einer Matrix

$$\mathbf{V}_h = [\mathbf{v}_1, \dots, \mathbf{v}_M] \quad (3.13)$$

zusammengefasst, so repräsentiert

$$\mathbf{y}_h(n) = \mathbf{V}_h^T \mathbf{x} \quad (3.14)$$

die M Hauptkomponenten. Somit ist $\mathbf{y}_h(n)$, im quadratischen Sinne, die beste M -dimensionale Repräsentation der Daten. Mit den übrigen Vektoren

$$\mathbf{V}_n = [\mathbf{v}_{M+1}, \dots, \mathbf{v}_N] \quad (3.15)$$

wird eine Beschreibung der Nebenkompenten

$$\mathbf{y}_n(n) = \mathbf{V}_n^T \mathbf{x}. \quad (3.16)$$

erreicht.

Jetzt kann damit die Anzahl der Quellen einer Mischung bestimmt werden. Existieren genau so viele Quellen wie Signale, so nehmen die Beträge der Eigenwerte $|\lambda_i|$ ungefähr gleiche Werte an. Dies gilt für den Fall, dass die einzelnen Quellen auch ungefähr gleich stark zu den gemischten Signalen beitragen.

Ganz anders sieht die Situation aus, wenn mehr gemischte Signale als Quellen vorliegen. Da dadurch auch weniger Hauptkomponenten existieren, tragen einzelne Dimensionen nicht zur Beschreibung der gemischten Daten bei. Sie enthalten lediglich ein eventuell vorhandenes Rauschen. Das spiegelt sich dann direkt in den Eigenwerten wider. Es gibt einige Eigenwerte, die betragsmäßig ungefähr gleich groß sind. Diese repräsentieren die echten Mischungen der Quellen. Die

3 Independent Component Analysis

restlichen Eigenwerte sind dann null oder nahezu null. Die dazugehörigen Dimensionen enthalten dann keine Daten oder nur ein, sowieso ungewolltes, Rauschen. Durch die Karhunen-Loeve-Transformation werden diese Komponenten dann entfernt. Die Anzahl der Dimensionen der übrig gebliebenen Daten ist auch gleichzeitig die Anzahl der Quellen.

3.1.3 Whitening

Eine weitere interessante Vorverarbeitungsstufe ist das sogenannte *Whitening* (Weißen) der Signale. Ein Signal ist weiß, wenn die Korrelationsmatrix die Identitätsmatrix ist:

$$E\{\mathbf{x}(n)\mathbf{x}^T(n)\} = \mathbf{I}. \quad (3.17)$$

Das ist eine stärkere Forderung als bei der PCA, da dort nur eine diagonale Korrelationsmatrix gefordert wird. Dies ist aber genau der Ausgangsschritt für das Whitening.

Bestimmung der Whiteningmatrix

Eine Whiteningmatrix $\mathbf{\Gamma}$ kann mit Hilfe der Ergebnisse der PCA Transformation bestimmt werden:

$$\mathbf{\Gamma} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{V}^T. \quad (3.18)$$

Hierbei sind $\mathbf{\Lambda}$ und \mathbf{V} wie in den Gleichungen (3.6) und (3.7) definiert. Die Dekorrelation und das Whitening der Signale kann mittels

$$\mathbf{y}(n) = \mathbf{\Gamma}\mathbf{x}(n) \quad (3.19)$$

erreicht werden. Ausgehend von der Annahme, dass $\mathbf{x}(n)$ so viele Komponenten wie Dimensionen hat, ist die Korrelationsmatrix \mathbf{R}_{xx} positiv definit, so dass alle Eigenwerte positiv sind:

$$\lambda_i > 0. \quad (3.20)$$

Damit kann $\mathbf{\Lambda}^{-\frac{1}{2}}$ als

$$\mathbf{\Lambda}^{-\frac{1}{2}} = \text{diag} \left\{ \frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \dots, \frac{1}{\sqrt{\lambda_N}} \right\} \quad (3.21)$$

berechnet werden. Man kann einfach zeigen, dass $\mathbf{\Gamma}$ eine Whitening-Matrix ist:

$$\mathbf{R}_{yy} = \mathbf{\Gamma}\mathbf{R}_{xx}\mathbf{\Gamma}^T = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\mathbf{V}\mathbf{\Lambda}^{-\frac{1}{2}} = \mathbf{I} \quad (3.22)$$

3.2 ICA basierend auf der Maximierung der Kurtosis

Hierbei wurde auf die Gleichung (3.5) und die Orthogonalität von \mathbf{V} mit

$$\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I} \quad (3.23)$$

zurückgegriffen.

$\mathbf{\Gamma}$ ist nicht die einzige Whitening-Matrix, denn mit jeder weiteren orthonormalen Matrix \mathbf{U} ist auch $\mathbf{\Gamma}' = \mathbf{U}\mathbf{\Gamma}$ eine Whitening-Matrix:

$$\mathbf{R}_{yy} = \mathbf{U}\mathbf{\Gamma}\mathbf{R}_{xx}\mathbf{\Gamma}^T\mathbf{U}^T = \mathbf{U}\mathbf{I}\mathbf{U}^T = \mathbf{I} \quad (3.24)$$

Dies zeigt noch deutlicher, dass alleine mit Statistik zweiter Ordnung eine Unabhängigkeit nicht erreicht werden kann.

Beispiel

Ein Beispiel für das Whitening-Verfahren ist in Abbildung 3.4 dargestellt. Der Ausgangspunkt ist eine beliebige Mischung zweier Signale, wie in Abbildung 3.4 (a) dargestellt. In Abbildung 3.4 (b) wurde das PCA Verfahren entsprechend Gleichung (3.3) angewandt. Die Signale wurden so gedreht, dass die Hauptkomponenten entlang der Achsen ausgerichtet sind.

Der entscheidende Schritt ist nun in Abbildung 3.4 (c) dargestellt. Hier wurden die dekorrelierten Daten so normiert, dass alle Komponenten die Energie eins haben. Durch diesen Schritt sind die gesuchten Komponenten orthogonal zueinander. Das heißt, dass unabhängige Signale durch eine orthonormale Drehung erreicht werden können, wie in Abbildung 3.4 (d) dargestellt.

Das Prinzip, unabhängige Komponenten mittels einer Drehung zu gewinnen, ist eine entscheidende Vereinfachung gegenüber der direkten Bestimmung einer Trennmatrix \mathbf{B} . Auf direktem Weg müssten insgesamt N^2 Einträge von \mathbf{B} bestimmt werden. Nach dem Whitening ist hingegen nur eine orthonormale Rotationsmatrix zu bestimmen, die aber nur $\frac{n(n-1)}{2}$ Freiheitsgrade besitzt. Für einen zweidimensionalen Fall bedeutet dies zum Beispiel, dass die Entmischungsmatrix für weiß gemachte Signale durch einen einzigen Parameter bestimmt ist.

3.2 ICA basierend auf der Maximierung der Kurtosis

In diesem Abschnitt wird nun eine Methode vorgestellt, die in der Lage ist, unabhängige Komponenten zu finden. Die grundsätzliche Überlegung stammt

3 Independent Component Analysis

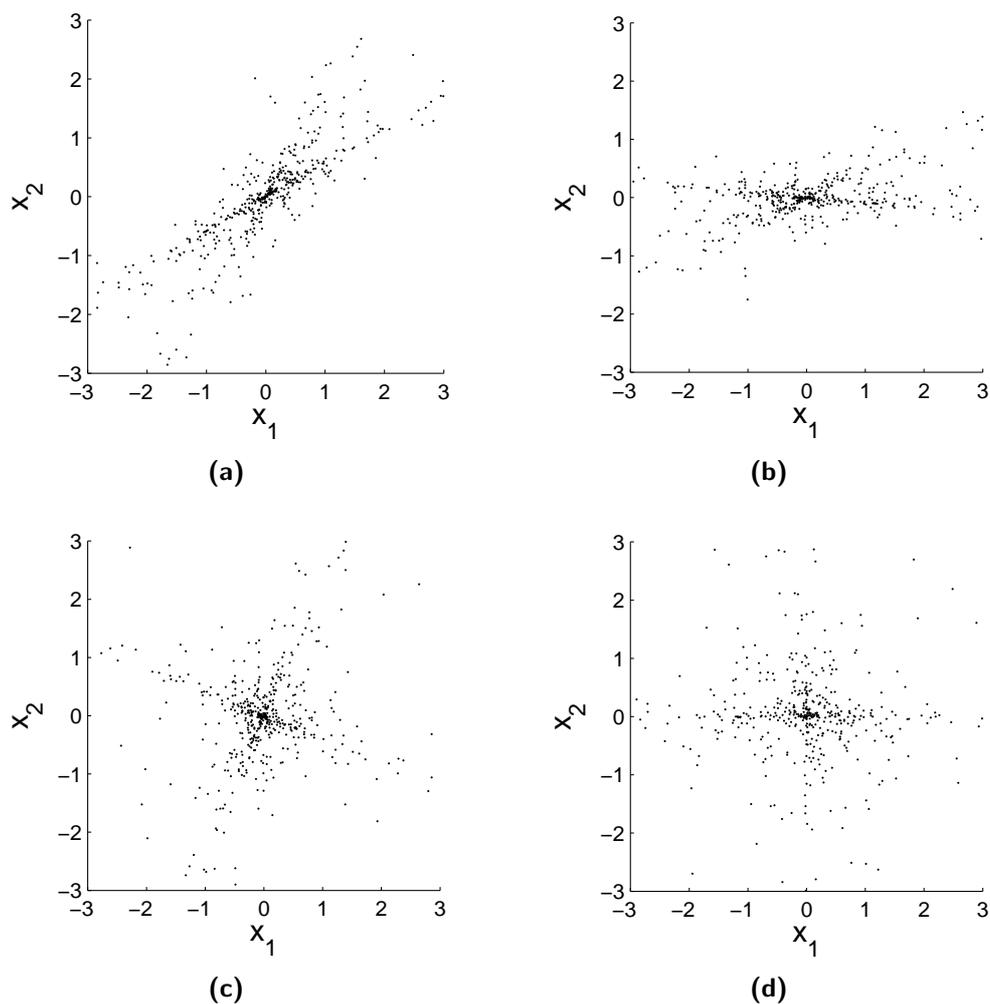


Abbildung 3.4: Das Whitening-Verfahren. (a) Die Mischung zweier Signale. (b) Dekorrelation mittels der PCA nach Gleichung (3.3). (c) Die Signale nach dem Whitening nach (3.19). (d) Unabhängige Signale können nach dem Whitening mittels einer einfachen Drehung gewonnen werden.

aus dem zentralen Grenzwertsatz. Es existieren mehrere Formulierungen. Eine davon besagt, dass die Wahrscheinlichkeitsverteilung einer unendlichen Summe von unabhängigen Variablen eine Gaußverteilung ist. Vereinfacht gesagt hat eine Summe von unabhängigen Variablen eine mehr gaußsche Verteilung als die einzelnen Summanden.

Dies wird nun zum Prinzip erhoben:

Die Suche nach unabhängigen Komponenten ist eine Suche nach nicht-gaußschen Komponenten.

Beispiel

Dieses Prinzip kann sehr eindrucksvoll anhand der Abbildungen 3.5 und 3.6 veranschaulicht werden. In Abbildung 3.5 (a) ist eine Mischung von zwei super-gaußschen Signalen als Streubild dargestellt. Daneben sind die beiden Randverteilungen dieser Mischung abgebildet. Zum Vergleich ist zusätzlich die Gaußverteilung eingezeichnet. In Abbildung 3.5 (b) sind die gleichen Signale nach der Trennung dargestellt. Zwar sind die unabhängigen Komponenten in den einzelnen Streubildern nicht gut erkennbar, aber der Unterschied bei den Randverteilungsdichten ist deutlich sichtbar. So haben unabhängige Signale eine größere Anzahl kleiner Werte als die gemischten Signale, was sich in den größeren Maxima um den Nullpunkt widerspiegelt. Die Verteilungsdichten der gemischten Signale sind der Gaußverteilung deutlich ähnlicher.

Ein ähnliches Bild kann auch für den Fall von subgaußschen Signalen aus Abbildung 3.6 beobachtet werden. Die Randverteilungsdichten der gemischten Signale in Abbildung 3.6 (a) nähern sich deutlicher an die Gaußverteilung an als die Randverteilungsdichten der getrennten Signale in Abbildung 3.6 (b).

In beiden Beispielen haben die gemischten Signale eine Verteilungsdichte, die näher an der Gaußverteilung ist als für die getrennten Signale. Ein Maß für diese Ähnlichkeit, die Kurtosis, wird im nächsten Abschnitt vorgestellt.

Zentrierung der Variablen

Die meisten Algorithmen für die ICA funktionieren nur bei mittelwertfreien Signalen. Das kann, wie schon bei der PCA, durch einfache Subtraktion des Mittelwertes erreicht werden:

$$\tilde{\mathbf{x}}(n) = \mathbf{x}(n) - E\{\mathbf{x}(n)\}. \quad (3.25)$$

Auf diese Art werden auch die getrennten Signale mittelwertfrei gemacht, denn eine lineare Abbildung verändert den Mittelwert nicht. Mit der Entmischungsmatrix \mathbf{B} ergeben sich dann die mittelwertfreien Quellen als

$$\tilde{\mathbf{s}}(n) = \mathbf{B}\tilde{\mathbf{x}}(n). \quad (3.26)$$

3 Independent Component Analysis

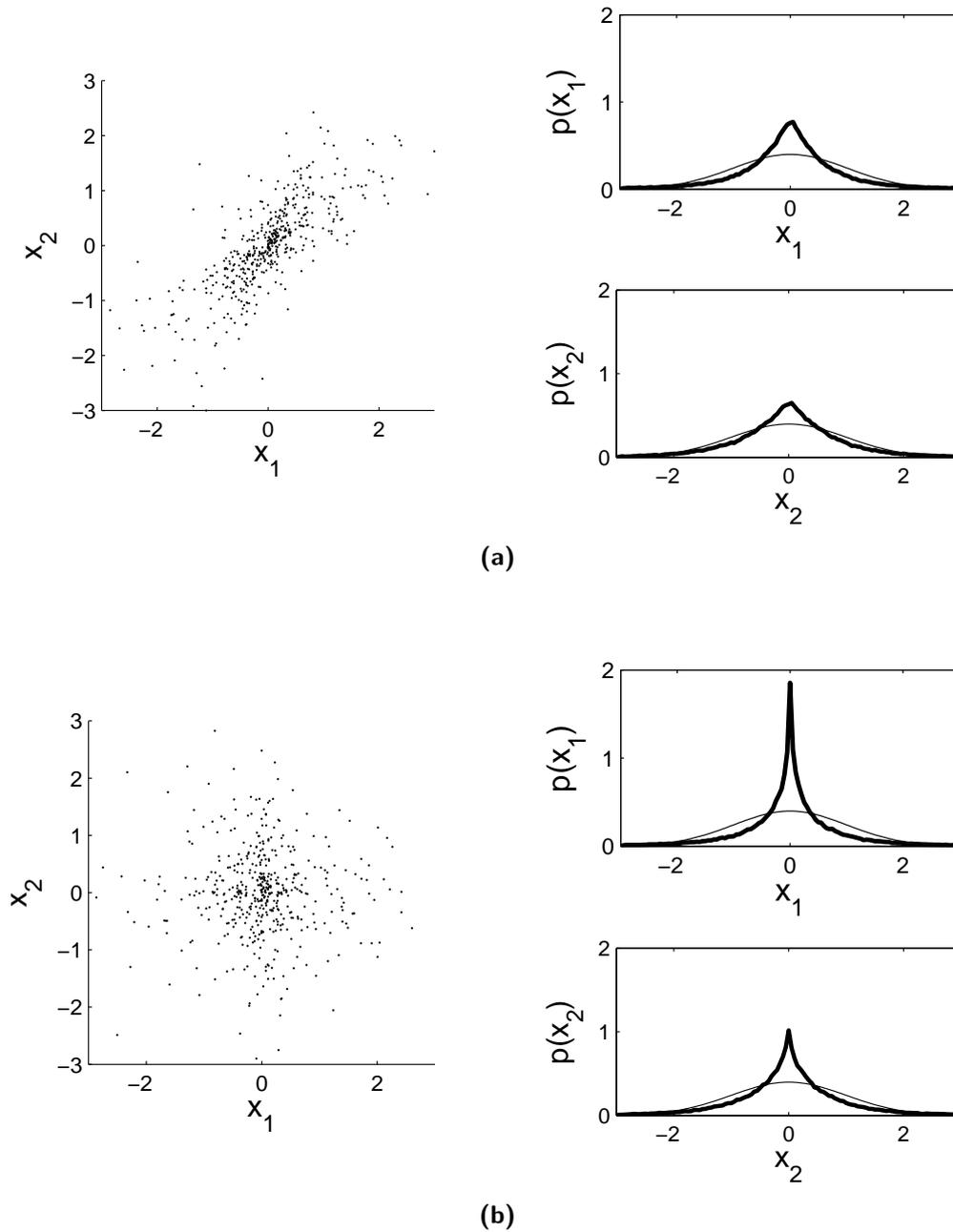


Abbildung 3.5: Beispiel für Randverteilungen von supergaußschen Signalen. (a) Zwei gemischte Signale und die dazugehörigen Randverteilungen. (b) Zwei getrennte Signale und die dazugehörigen Randverteilungen.

3.2 ICA basierend auf der Maximierung der Kurtosis

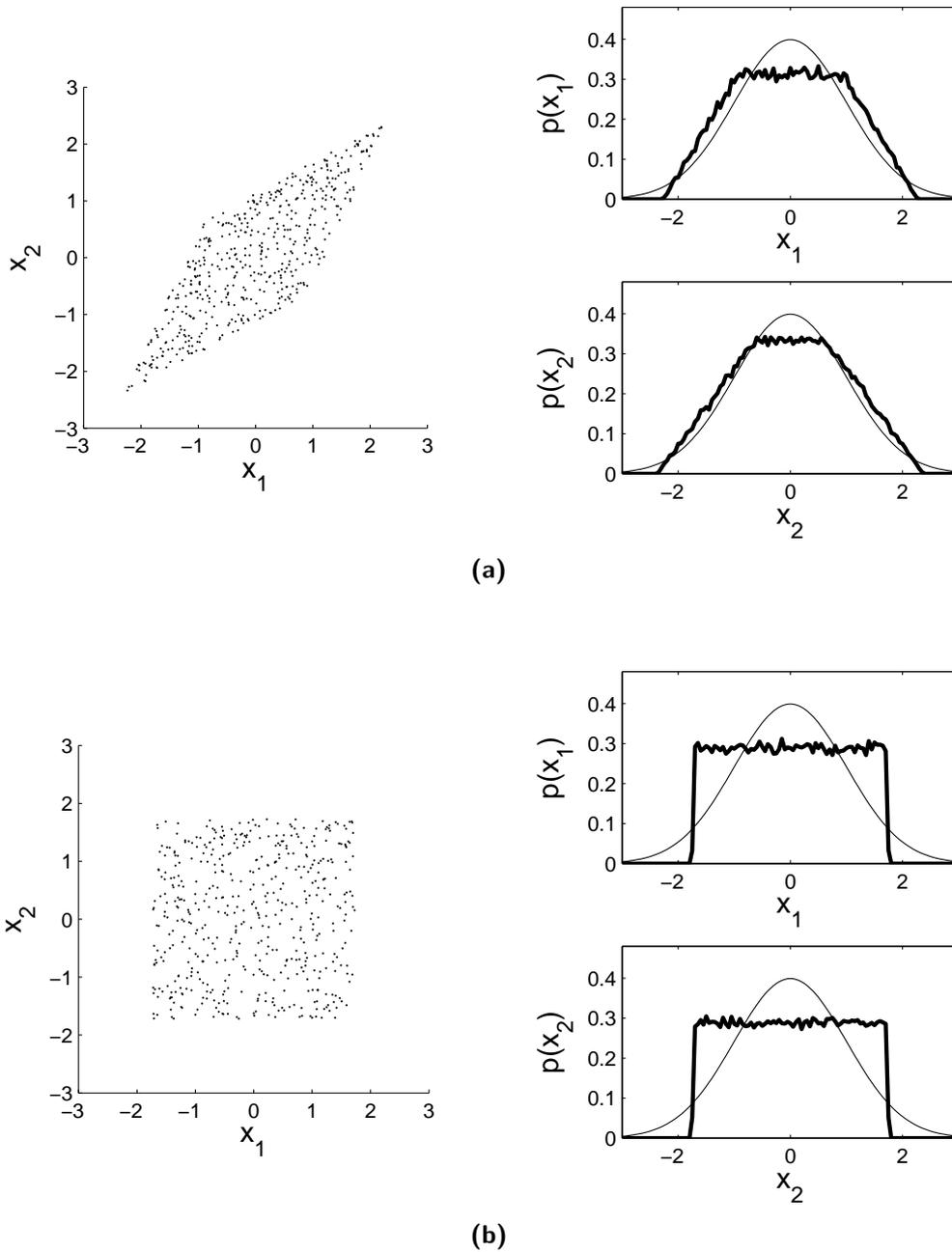


Abbildung 3.6: Beispiel für Randverteilungen von subgaußschen Signalen. (a) Zwei gemischte Signale und die dazugehörigen Randverteilungen. (b) Zwei getrennte Signale und die dazugehörigen Randverteilungen.

3 Independent Component Analysis

Die Bestimmung der Entmischungsmatrix wird so einfacher, und falls der Offset gewünscht ist, kann er nachträglich mit $\mathbf{B}E\{\mathbf{x}(n)\}$ addiert werden.

Im Folgenden wird davon ausgegangen, dass alle Signale mittelwertfrei sind. Dies kann entweder mit Hilfe von Gleichung (3.26) oder der Whitening-Transformation durchgeführt werden.

3.2.1 Kurtosis als Maß für die Gaußheit

Ein bekanntes Maß, mit dem die Gaußheit einer Verteilung gemessen werden kann, ist die Kurtosis. Sie wird mit Hilfe des zweiten und vierten Moments definiert:

$$\text{kurt}(x) = E\{x^4\} - 3E\{x^2\}^2. \quad (3.27)$$

Mit dieser Definition gilt:

$$\text{kurt}(x_1 + x_2) = \text{kurt}(x_1) + \text{kurt}(x_2) \quad (3.28)$$

und

$$\text{kurt}(\alpha x_1) = \alpha^4 \text{kurt}(x_1). \quad (3.29)$$

Mit dieser Formulierung bekommt der Wert der Kurtosis folgende Bedeutung: Für den Fall, dass die Kurtosis den Wert null hat, handelt es sich um eine Gaußverteilung. Für positive Werte handelt es sich um eine subgaußsche Verteilung wie beispielsweise die Gleichverteilung. Bei negativen Werten ist es eine supergaußsche Verteilung. Ein Beispiel dafür ist die Laplaceverteilung.

Das Prinzip der Suche nach nichtgaußschen Komponenten ist mithin eine Suche nach Komponenten, die eine betragsmäßig größtmögliche Kurtosis haben.

Beispiel

Der Zusammenhang zwischen der Kurtosis und den unabhängigen Komponenten ist in Abbildung 3.7 dargestellt. In Abbildung 3.7 (a) ist das Streubild einer Mischung mit einem Projektionsvektor gezeigt. Dieser Projektionsvektor kann auf dem Einheitskreis rotiert werden. In Abhängigkeit des Rotationswinkels α kann für jede Projektion die Kurtosis berechnet werden. Die Werte der Kurtosis sind in Abbildung 3.7 (b) dargestellt. Der Verlauf ist aufgrund der Symmetrie periodisch, so dass nur der Bereich $[0, \pi]$ interessant ist. In diesem Beispiel hat der Verlauf der Kurtosis bei ungefähr $5/6\pi$ und $7/8\pi$ jeweils ein lokales Maximum. Bei diesen Winkeln ist der Entmischungsvektor senkrecht zu einer

3.2 ICA basierend auf der Maximierung der Kurtosis

der Komponenten. Bei der Projektion mit diesen Winkeln bleibt jeweils nur die andere Komponente übrig.

Bei der Mischung von zwei subgaußschen Signalen aus Abbildung 3.8 ergibt sich ein ganz ähnliches Bild. Hier werden die unabhängigen Komponenten allerdings durch die beiden lokalen Minima bei den gleichen Winkeln bestimmt.

In beiden Beispielen können die unabhängigen Komponenten durch eine Projektion mit der betragsmäßig größten Kurtosis extrahiert werden.

3.2.2 Gradientenverfahren

Die Suche nach den Projektionen mit größter Kurtosis kann mit Hilfe eines Gradientenabstiegsverfahrens durchgeführt werden [30]. Es ist eine Suche nach einer Linearkombination \mathbf{b} der gemischten Signale, so dass die Kurtosis von $\mathbf{b}^T \mathbf{x}$ betragsmäßig maximal wird³. Mit

$$\mathbf{z}^T = \mathbf{b}^T \mathbf{A} \quad (3.30)$$

gilt für weiß gemachte Signale

$$\text{kurt}(\mathbf{b}^T \mathbf{x}) = \text{kurt}(\mathbf{b}^T \mathbf{A} \mathbf{s}) = \text{kurt}(\mathbf{z}^T \mathbf{s}) = \sum_{i=0}^{N-1} z_i^4 \text{kurt}(s_i). \quad (3.31)$$

Mit der Einschränkung der Misch- und Entmischungsvektoren auf eine konstante Länge, $\|\mathbf{b}\| = \|\mathbf{z}\| = 1$, wurde in [23] gezeigt, dass die Extrema von (3.31) genau die gesuchten unabhängigen Komponenten sind.

Da die Eingangssignale von (3.31) weiß sind, gilt:

$$\text{kurt}(\mathbf{b}^T \mathbf{x}) = E\{(\mathbf{b}^T \mathbf{x})^4\} - 3E\{(\mathbf{b}^T \mathbf{x})^2\}^2 = E\{(\mathbf{b}^T \mathbf{x})^4\} - 3\|\mathbf{b}\|^4. \quad (3.32)$$

Da der absolute Betrag der Kurtosis maximiert werden soll, lautet die Kostenfunktion also

$$J(\mathbf{b}) = |E\{(\mathbf{b}^T \mathbf{x})^4\} - 3\|\mathbf{b}\|^4|. \quad (3.33)$$

Der Gradient bezüglich \mathbf{b} lautet

$$\frac{\partial |\text{kurt}(\mathbf{b}^T \mathbf{x})|}{\partial \mathbf{b}} = \text{sgn}(\text{kurt}(\mathbf{b}^T \mathbf{x})) \cdot [E\{\mathbf{x}(\mathbf{b}^T \mathbf{x})^3\} - 3\|\mathbf{b}\|^2 \mathbf{b}]. \quad (3.34)$$

³Bei der Herleitung wird der Zeitindex n wegelassen; statt des Zufallssignals $\mathbf{x}(n)$ wird die Zufallsvariable \mathbf{x} benutzt.

3 Independent Component Analysis

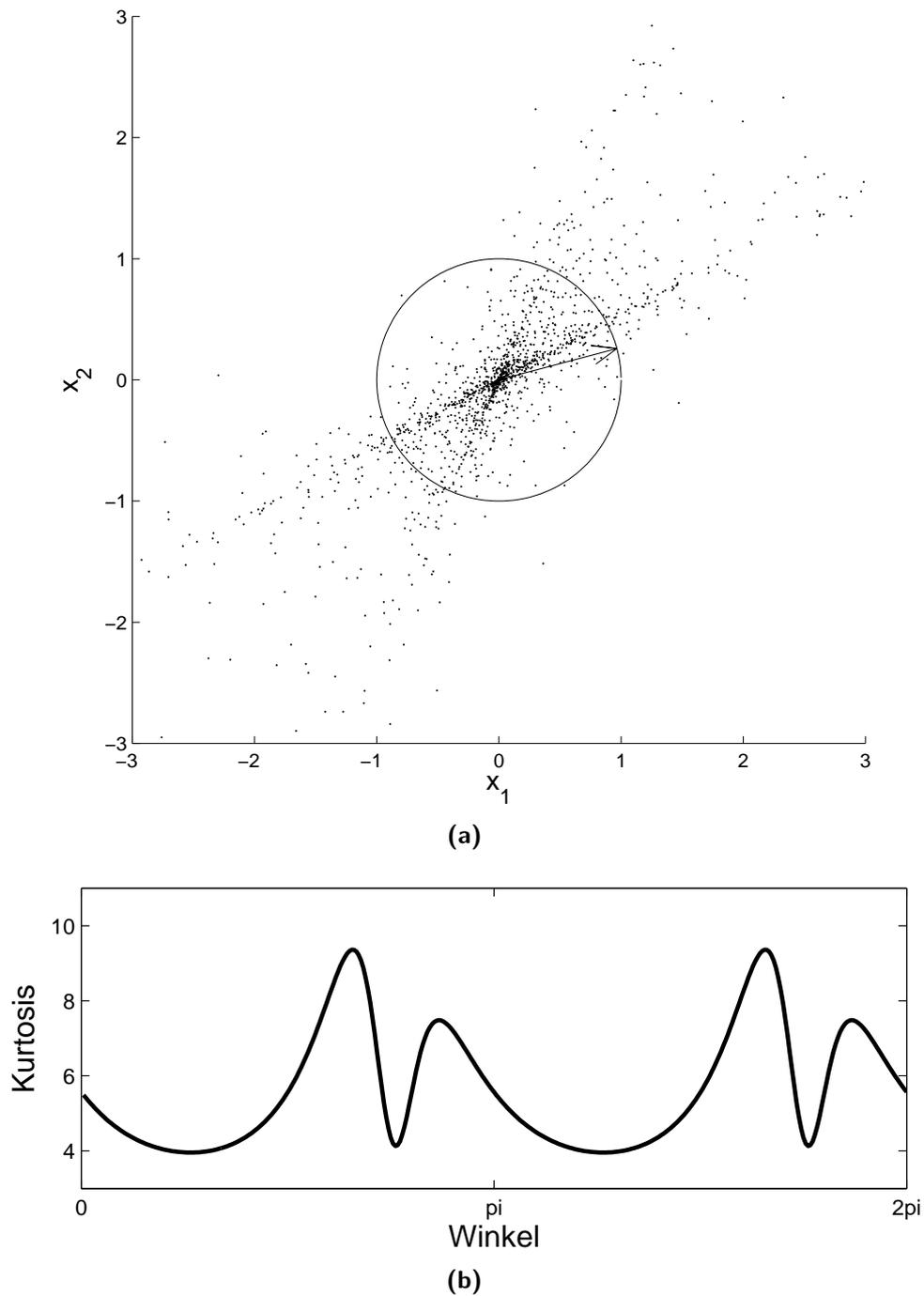
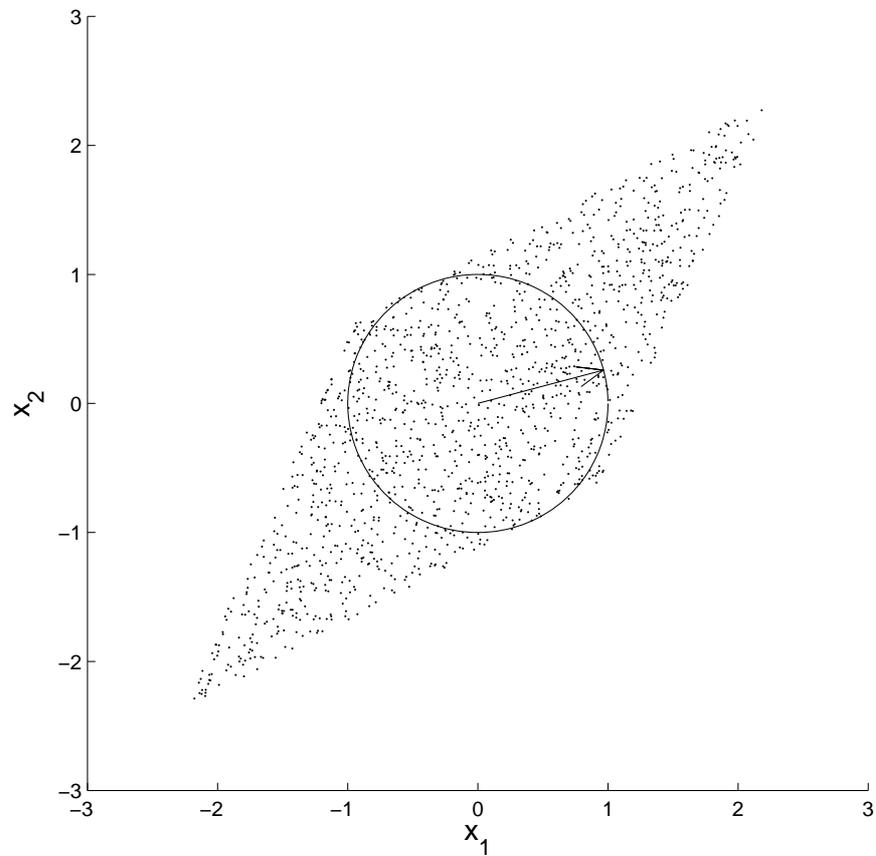
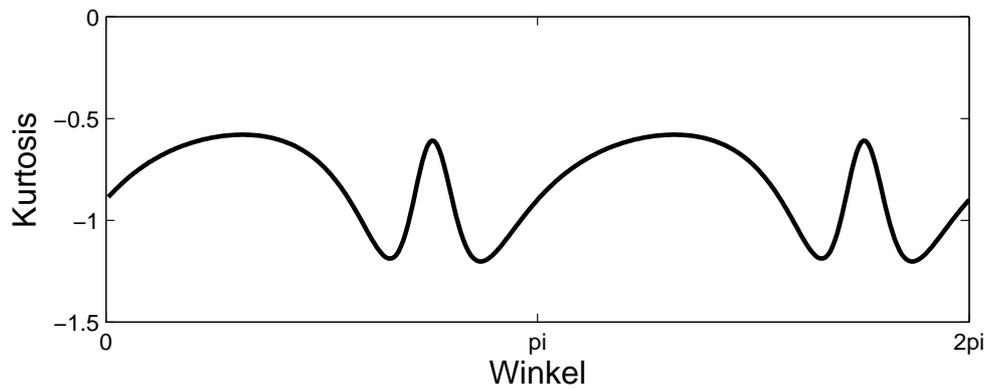


Abbildung 3.7: Beispiel für Abhängigkeit der Kurtosis vom Projektionsvektor. (a) Streubild zweier gemischter supergaußer Signale mit einem Projektionsvektor. (b) Die Kurtosis der projizierten Signale in Abhängigkeit des Winkels des Projektionsvektors.

3.2 ICA basierend auf der Maximierung der Kurtosis



(a)



(b)

Abbildung 3.8: Beispiel für Abhängigkeit der Kurtosis vom Projektionsvektor. (a) Streubild zweier gemischter subgaußscher Signale mit einem Projektionsvektor. (b) Die Kurtosis der projizierten Signale in Abhängigkeit des Winkels des Projektionsvektors.

3 Independent Component Analysis

Der erste Term bestimmt lediglich, ob die Signale super- oder subgaußisch sind. Sofern bekannt, wie zum Beispiel bei Sprachsignalen, kann der Term durch das entsprechende Vorzeichen ersetzt werden. Der letzte Term verändert nur die Länge, aber nicht die Richtung des Entmischungsvektors, und kann bei entsprechender Normierung weggelassen werden. Eine derartige Normierung ist ohnehin notwendig, da die Nebenbedingung der konstanten Länge des Entmischungsvektors \mathbf{b} in Gleichung (3.34) noch fehlt. Diese Nebenbedingung kann zum Beispiel durch eine Projektion von \mathbf{b} nach jedem Updateschritt auf den Einheitskreis erfüllt werden. Somit ergibt sich für Sprachsignale das folgende Gradientenverfahren:

$$\Delta \mathbf{b} = \mu \cdot E\{\mathbf{x}(\mathbf{b}^T \mathbf{x})^3\}, \quad (3.35)$$

$$\mathbf{b}^{l+1} = \frac{\mathbf{b}^l + \Delta \mathbf{b}}{\|\mathbf{b}^l + \Delta \mathbf{b}\|_{\ell_2}}. \quad (3.36)$$

3.2.3 Fixpunkt Algorithmus

Die Wahl der Lernrate μ aus Gleichung (3.35) ist entscheidend für die Geschwindigkeit der Konvergenz. Bei kleinen μ kann es lange dauern, bis eine Lösung gefunden ist. Bei zu großen μ besteht hingegen die Gefahr, dass der Algorithmus nicht konvergiert.

Eine Alternative ist eine Fixpunktiteration. Diese läßt sich aus der folgenden Beobachtung herleiten: In einem stabilen Punkt des Gradientenverfahrens muss der Gradient in Richtung des Entmischungsvektors \mathbf{b} zeigen. Das heißt: Der Gradient ist ein Vielfaches von \mathbf{b} :

$$\mathbf{b} \propto E\{\mathbf{x}(\mathbf{b}^T \mathbf{x})^3\} - 3\|\mathbf{b}\|^2 \mathbf{b}. \quad (3.37)$$

Mit dem gleichen Normalisierungsschritt ergibt sich das folgende Verfahren:

$$\mathbf{b}' = E\{\mathbf{x}(\mathbf{b}^T \mathbf{x})^3\} - 3\mathbf{b} \quad (3.38)$$

$$\mathbf{b}^{l+1} = \frac{\mathbf{b}^l + \mathbf{b}'}{\|\mathbf{b}^l + \mathbf{b}'\|_{\ell_2}}. \quad (3.39)$$

Dieser Algorithmus wurde in [30] präsentiert und ist bekannt unter dem Namen FastICA. Das Besondere an diesem Verfahren ist, dass es ohne Lernrate oder weitere Parameter auskommt. Weiterhin ist die Konvergenzgeschwindigkeit, wie in [30] gezeigt, kubisch. Insgesamt kann es im Allgemeinen als schneller und stabiler als das Gradientenverfahren bewertet werden.

Extraktion mehrerer Quellen

In der hier vorgestellten Form extrahiert der Algorithmus aus Abschnitt 3.2.3 genau eine Quelle. Will man mehrere oder gar alle Quellen bestimmen, bieten sich mehrere Möglichkeiten.

- Der Algorithmus wird an zufälligen Startpunkten gestartet. Je nach Startpunkt ergibt sich ein anderer Konvergenzpunkt und somit eine andere Quelle. Es existiert keine Möglichkeit, sicher festzustellen, ob auch alle Quellen gefunden wurden.
- Eine alternative Methode ist die *sequentielle Extraktion* [19, 75]. Die zentrale Idee ist es, die einzeln bestimmten Signale $y_j(n)$ aus der Mischung $\mathbf{x}^j(n)$ zu entfernen und das Verfahren zu wiederholen, bis alle Signale getrennt vorliegen. Dazu wird mit einem Entmischungsvektor \mathbf{b}_j ein Signal $y_j(n)$ extrahiert. Mit diesem Wissen wird nun die gefundene Quelle aus der Mischung \mathbf{x}^j entfernt:

$$\mathbf{x}^{j+1}(n) = \mathbf{x}^j(n) - \mathbf{b}_j y_j(n). \quad (3.40)$$

Auf die neuen Signale $\mathbf{x}^{j+1}(n)$ kann wieder die FastICA angewendet und eine weitere Quelle $y_{j+1}(n)$ bestimmt werden. Diese Prozedur wird so lange wiederholt, bis alle Quellen gefunden worden sind.

Auch hier liegt die Schwierigkeit darin, insbesondere bei vorhandenem Rauschen, dass es unklar sein kann, wann die letzte Quelle bestimmt worden ist und das Verfahren abgebrochen werden kann. Weiterhin existiert das Problem, dass bei einer nicht hinreichend guten Bestimmung eines Entmischungsvektors eine Quelle nicht vollständig entfernt wird. Mithin besteht das Risiko, anstatt einer echten weiteren, dieselbe Quelle noch einmal zu finden.

- Es kann auch gleichzeitig nach allen Quellen gesucht werden. Hierbei werden in jedem Iterationsschritt alle Entmischungsvektoren \mathbf{b}_j parallel aktualisiert und zu einer Entmischungsmatrix \mathbf{B} zusammengefasst. Wurden die Signale bei der Vorverarbeitung weiß gemacht, so muss die Entmischungsmatrix \mathbf{B} orthonormal sein. Dies kann nun ausgenutzt werden, indem in jedem Schritt \mathbf{B} auf die Menge der orthonormalen Matrizen projiziert wird. Somit können alle Entmischungsvektoren gleichzeitig bestimmt werden.

3.2.4 Alternativen zur Messung der Gaußheit

Der FastICA Algorithmus, wie im letzten Abschnitt beschrieben, benutzt zur Messung der Gaußheit die Kurtosis. Die Bestimmung der Kurtosis ist bei einer endlichen Stichprobe im Allgemeinen sehr fehleranfällig. So reichen bereits einzelne Ausreißer, um den Wert der Kurtosis erheblich zu verfälschen. Diese Fehler basieren insbesondere auf der vierten Potenz des vierten Moments aus Gleichung (3.27). So reicht beispielsweise bei einer Normalverteilung mit der Kurtosis null ein einzelner Ausreißer des Werts 10, um die Kurtosis auf einen Wert auf über sieben zu verfälschen. Insgesamt ist diese Methode nicht sehr robust.

Negentropie

Eine weitere Methode zur Messung der Gaußheit ist die differenzielle Entropie beziehungsweise die darauf aufbauende Negentropie. Die differenzielle Entropie ist ein Maß für die Zufälligkeit einer Variablen. Je zufälliger, also unvorhersagbarer eine Variable ist, umso größer ist ihre Entropie. Die differenzielle Entropie H eines Zufallsvektors \mathbf{y} mit der Verteilungsdichte $p_{\mathbf{y}}(\zeta)$ ist definiert als

$$H(\mathbf{y}) = - \int_{\zeta} p_{\mathbf{y}}(\zeta) \log p_{\mathbf{y}}(\zeta) d\zeta. \quad (3.41)$$

$H(\mathbf{y})$ ist maximal, wenn es sich bei $p_{\mathbf{y}}(\zeta)$ um eine gaußsche Verteilung handelt. Das bedeutet für eine gaußsche Variable, im Vergleich zu allen anderen Variablen mit der gleichen Kovarianzmatrix, die größte Entropie beziehungsweise die größte Zufälligkeit. Das kann man nun ausnutzen, um ein Maß zu erhalten, welches null für eine gaußsche Variable liefert und nichtnegativ ist. Hierzu wird die Negentropie als der Abstand zur einer Gaußverteilung definiert:

$$J(\mathbf{y}) = H(\mathbf{y}_{\text{gauss}}) - H(\mathbf{y}). \quad (3.42)$$

Hierbei ist $\mathbf{y}_{\text{gauss}}$ eine Zufallsvariable mit einer Normalverteilung gleicher Kovarianzmatrix und Mittelwert wie \mathbf{y} .

Die Suche nach unabhängigen Quellen wird auf diese Weise zur Suche nach Komponenten, die eine möglichst große Negentropie haben [27]. Die Negentropie ist zwar ein gutes Maß für die Gaußheit einer Verteilung [21, 22, 14], jedoch ist die praktische Umsetzung schwierig. Es bedarf insbesondere der Bestimmung der Verteilungsdichte. Diese ist oftmals unbekannt und möglicherweise sogar parametrisch nicht zu beschreiben. Alternativ können auch Näherungen benutzt

3.3 ICA basierend auf der Kullback-Leibler-Divergenz

werden, diese sind dann oftmals nicht robuster als die bereits besprochene Kurtosis.

Robuste Nichtlinearitäten

Eine weitere Alternative ist die Benutzung der sogenannten nichtlinearen Momente [29]. Hierbei wird das vierte Moment aus der Gleichung (3.33) durch eine andere Funktion G_i ersetzt. Durch eine geschickte Wahl von G_i kann eine wesentlich bessere Robustheit erreicht werden. Dies gilt insbesondere für Funktionen, die auf einzelne Ausreißer nicht so empfindlich reagieren, also für große Werte nicht so stark ansteigen. Mögliche Beispiele [29] sind:

$$G_1(y) = \frac{1}{a_1} \log \cosh(a_1 y), \quad (3.43)$$

$$G_2(y) = -\exp\left(-\frac{y^2}{2}\right), \quad (3.44)$$

$$G_3(y) = y^4. \quad (3.45)$$

Die dritte Funktion ist die bereits bekannte Kurtosis und ist zum Vergleich angegeben. Für das Gradientenverfahren aus Gleichung (3.35) ergeben sich damit die folgenden Funktionen:

$$g_1(y) = \tanh(a_1 y), \quad (3.46)$$

$$g_2(y) = y \exp\left(-\frac{y^2}{2}\right), \quad (3.47)$$

$$g_3(y) = y^3. \quad (3.48)$$

Die robusten Nichtlinearitäten $G_i(y)$ und die Ableitungen $g_i(y)$ sind in Abbildung 3.9 dargestellt.

3.3 ICA basierend auf der Kullback-Leibler-Divergenz

In diesem Abschnitt wird eine weitere Methode für die ICA vorgestellt. Sie nutzt die Nichtgaußheit der Quellen auf andere Art und Weise und vermeidet die Nachteile der Kurtosis-basierten Verfahren.

3 Independent Component Analysis

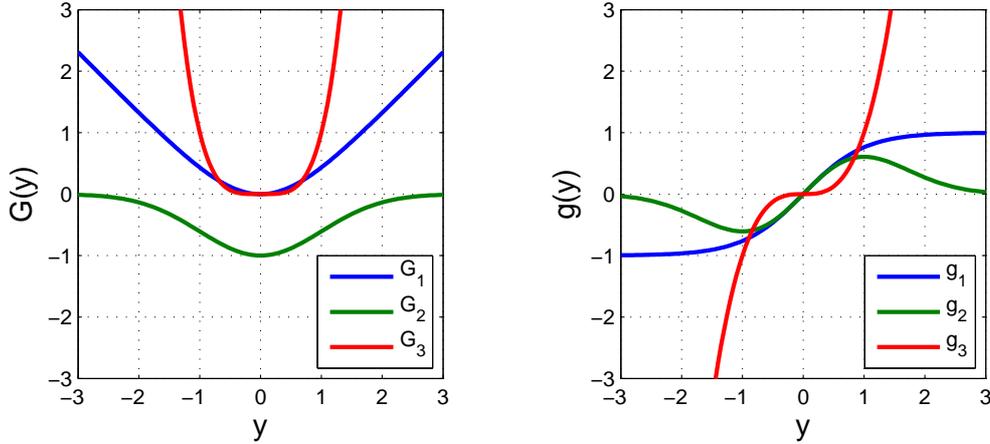


Abbildung 3.9: Die robusten Nichtlinearitäten aus den Gleichungen (3.43) bis (3.45) und die dazugehörigen Ableitungen. $G_3(y)$ ist die Kurtosis, die zu Vergleichszwecken mitangeführt wird.

Die im Folgenden vorgestellte Methode basiert auf der Tatsache, dass die Verteilungsdichte $p_y(\mathbf{y})$ für unabhängige Komponenten in einzelne Randverteilungsdichten $p_i(y_i)$ mit

$$p_y(\mathbf{y}) = \prod_{i=1}^N p_i(y_i) \quad (3.49)$$

separiert werden kann. Diese Tatsache wird nun genutzt, um eine Kostenfunktion zu definieren, die angibt, wie stark sich die Verteilungsdichte der aufgenommenen Signale von einer Verteilungsdichte unterscheidet, die sich separieren lässt.

Das Bilden eines Gradienten bezüglich dieser Kostenfunktion führt dann zu einem sehr robusten Algorithmus für die ICA.

3.3.1 Definitionen

Verteilungsdichten Für eine mehrdimensionale Verteilungsdichte $p_y(\mathbf{y})$ wird die i -te Randverteilungsdichte $p_i(y_i)$ berechnet, indem über alle anderen Dimensionen integriert wird:

$$p_i(y_i) = \int_{-\infty}^{\infty} p_y(\mathbf{y}) d\check{\mathbf{y}}^i. \quad (3.50)$$

Hierbei ist

$$\check{\mathbf{y}}^i = [y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n]^T \quad (3.51)$$

3.3 ICA basierend auf der Kullback-Leibler-Divergenz

der Vektor, der übrig bleibt, wenn die i -te Komponente aus \mathbf{y} entfernt wird. Aus der Definition folgt ebenfalls:

$$d\mathbf{y} = d\check{\mathbf{y}}^i dy_i. \quad (3.52)$$

Differenzielle Entropie Die differenzielle Entropie einer Zufallsvariablen \mathbf{y} ist wie bereits in (3.41) als

$$H(\mathbf{y}) = - \int_{-\infty}^{\infty} p_y(\mathbf{y}) \log p_y(\mathbf{y}) d\mathbf{y} = -E\{\log p_y(\mathbf{y})\} \quad (3.53)$$

definiert. Eine lineare Abbildung

$$\mathbf{y} = \mathbf{B}\mathbf{x} \quad (3.54)$$

verändert die differenzielle Entropie [66]:

$$H(\mathbf{y}) = H(\mathbf{x}) + \log |\det(\mathbf{B})|. \quad (3.55)$$

Kullback-Leibler-Divergenz Die Kullback-Leibler-Divergenz (KLD) K_{pq} ist ein Maß dafür, wie stark sich zwei Verteilungsdichten $p_y(\mathbf{y})$ und $q_y(\mathbf{y})$ unterscheiden:

$$K_{pq} = K[p_y(\mathbf{y})||q_y(\mathbf{y})] = \int_{-\infty}^{\infty} p_y(\mathbf{y}) \log \frac{p_y(\mathbf{y})}{q_y(\mathbf{y})} d\mathbf{y}. \quad (3.56)$$

Die Kullback-Leibler-Divergenz erfüllt nicht alle Anforderungen an eine Metrik, da die Symmetrie nicht gegeben ist. Trotzdem ist sie ein geeignetes Maß, da sie immer positiv ist und nur für ein identisches $p_y(\mathbf{y})$ und $q_y(\mathbf{y})$ gleich null ist.

Kostenfunktion Ausgehend von Gleichung (3.54) wird nun eine sinnvolle Kostenfunktion definiert. Sie soll angeben, wie unabhängig die Verteilungsdichte $p_y(\mathbf{y})$ der mittels \mathbf{B} transformierten Signale ist. Dies wird erreicht, indem $p_y(\mathbf{y})$ mit dem Produkt der Randverteilungen $p_i(y_i)$ mittels der Kullback-Leibler-Divergenz verglichen wird:

$$\begin{aligned} \mathcal{R}(\mathbf{B}, \mathbf{y}) &= K \left[p_y(\mathbf{y}) \left\| \prod_{i=1}^N p_i(y_i) \right\| \right] \\ &= \int_{-\infty}^{\infty} p_y(\mathbf{y}) \log \frac{p_y(\mathbf{y})}{\prod_{i=1}^N p_i(y_i)} d\mathbf{y}. \end{aligned} \quad (3.57)$$

3 Independent Component Analysis

Die Kostenfunktion $\mathcal{R}(\mathbf{B}, \mathbf{y})$ läßt sich auch mit Hilfe der differentiellen Entropie ausdrücken:

$$\begin{aligned}
 \mathcal{R}(\mathbf{B}, \mathbf{y}) &= \int_{-\infty}^{\infty} p_{\mathbf{y}}(\mathbf{y}) \log \frac{p_{\mathbf{y}}(\mathbf{y})}{\prod_{i=1}^N p_i(y_i)} d\mathbf{y} \\
 &= \int_{-\infty}^{\infty} p_{\mathbf{y}}(\mathbf{y}) \log p_{\mathbf{y}}(\mathbf{y}) d\mathbf{y} - \int_{-\infty}^{\infty} p_{\mathbf{y}}(\mathbf{y}) \log \prod_{i=1}^N p_i(y_i) d\mathbf{y} \\
 &= -H(\mathbf{y}) - \sum_{i=1}^N \int_{-\infty}^{\infty} p_{\mathbf{y}}(\mathbf{y}) \log p_i(y_i) d\mathbf{y} \\
 &= -H(\mathbf{y}) + \sum_{i=1}^N H_i(y_i).
 \end{aligned} \tag{3.58}$$

Für den letzten Schritt wurden die Eigenschaften der Randverteilungen genutzt. Zusammen mit (3.52) folgt

$$\begin{aligned}
 \int_{-\infty}^{\infty} p_{\mathbf{y}}(\mathbf{y}) \log p_i(y_i) d\mathbf{y} &= \int_{-\infty}^{\infty} \log p_i(y_i) \int_{-\infty}^{\infty} p_{\mathbf{y}}(\mathbf{y}) d\mathbf{y}^i dy_i \\
 &= \int_{-\infty}^{\infty} p_i(y_i) \log p_i(y_i) dy_i \\
 &= -H_i(y_i).
 \end{aligned} \tag{3.59}$$

Die Kostenfunktion aus Gleichung (3.58) kann noch weiter umformuliert werden. Mit der Gleichung (3.55) und der Definition (3.53) ist

$$\begin{aligned}
 \mathcal{R}(\mathbf{B}, \mathbf{y}) &= -H(\mathbf{y}) + \sum_{i=1}^N H_i(y_i) \\
 &= -H(\mathbf{x}) - \log |\det(\mathbf{B})| - \sum_{i=1}^N E\{\log p_i(y_i)\}.
 \end{aligned} \tag{3.60}$$

Bei dieser Formulierung ist die Kostenfunktion unabhängig von der differentiellen Entropie $H(\mathbf{x})$ der Eingangssignale $\mathbf{x}(n)$. Sofern die Kostenfunktion lediglich für eine Bestimmung von \mathbf{B} benutzt wird, kann dieser Term auch weggelassen werden.

Beispiel

Ein Beispiel für die Kullback-Leibler-Divergenz und die einzelnen Schritte der Herleitung der Kostenfunktion ist in den Abbildungen 3.10 und 3.11 gegeben. In Abbildung 3.10 (a) sind eine Verteilungsdichte $p_y(\mathbf{y})$ und die dazugehörigen Randverteilungen $p_i(y_i)$ dargestellt. In diesem Fall ist $p_y(\mathbf{y})$ mit

$$p_y(\mathbf{y}) = p_1(y_1) \cdot p_2(y_2) \quad (3.61)$$

separierbar. Damit sind die dazugehörigen Zufallsvariablen y_1 und y_2 voneinander unabhängig. Dies gilt nicht für die Verteilungsdichte aus Abbildung 3.10 (b). Hier ist $p_y(\mathbf{y})$ nicht separierbar. Das Produkt der Randverteilungen ist in Abbildung 3.11 (a) dargestellt. Die beiden Verteilungsdichten aus Abbildungen 3.10 (b) und 3.11 (a) unterscheiden sich deutlich voneinander. Dieser Unterschied wird mit Hilfe der KLD gemessen. Der dazu benutzte Quotient

$$Q_{\text{KLD}}(\mathbf{y}) = p_y(\mathbf{y}) \log \frac{p_y(\mathbf{y})}{\prod_{i=1}^N p_i(y_i)} \quad (3.62)$$

aus Gleichung (3.57) ist in Abbildung 3.11 (b) dargestellt.

3.3.2 Gradientenverfahren

Euklidischer Gradient Gesucht wird nun das Minimum der Kostenfunktion $\mathcal{R}(\mathbf{B}, \mathbf{y})$, weil die dazugehörige Matrix \mathbf{B} genau die gesuchte Entmischungsmatrix ist. Da das Minimum von $\mathcal{R}(\mathbf{B}, \mathbf{y})$ analytisch nicht bestimmbar ist, wird auch hier ein Gradientenabstiegsverfahren

$$\mathbf{B}^{l+1} = \mathbf{B}^l + \mu \Delta \mathbf{B} \quad (3.63)$$

mit

$$\Delta \mathbf{B} = -\frac{\partial \mathcal{R}(\mathbf{B}, \mathbf{y})}{\partial \mathbf{B}} \quad (3.64)$$

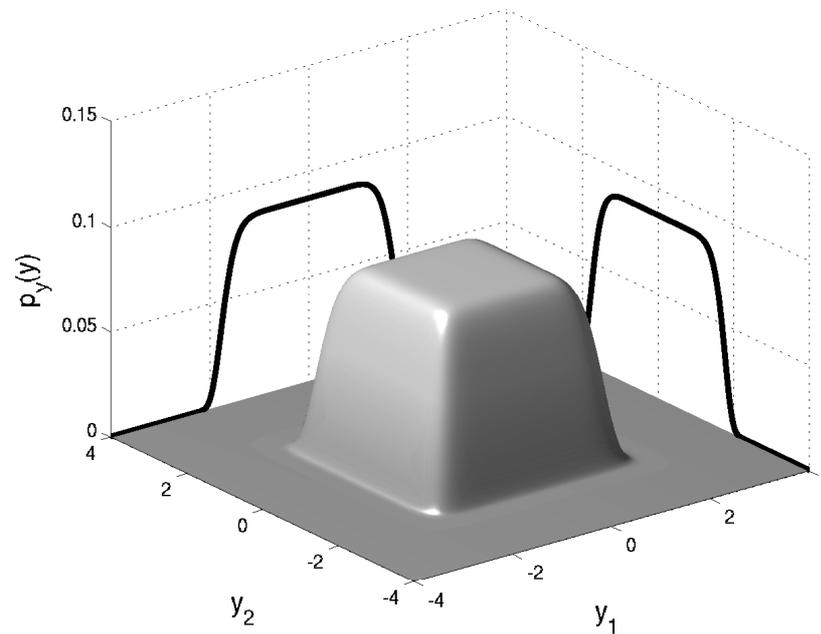
benutzt. Ausgehend von der vereinfachten Kostenfunktion

$$\mathcal{R}'(\mathbf{B}, \mathbf{y}) = -\log |\det(\mathbf{B})| - \sum_{i=1}^N E\{\log p_i(y_i)\} \quad (3.65)$$

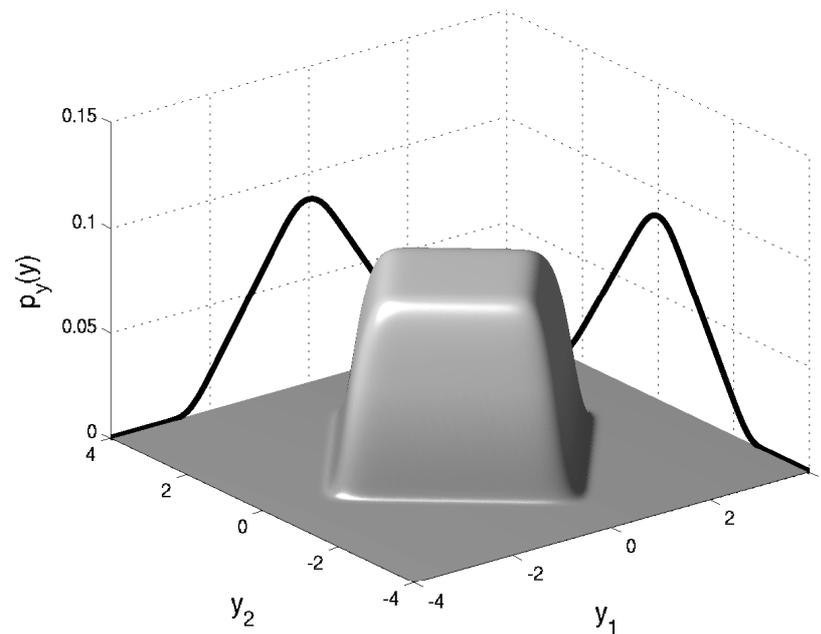
kann der Gradient

$$\Delta \mathbf{B} = -\frac{\partial \mathcal{R}(\mathbf{B}, \mathbf{y})}{\partial \mathbf{B}} = (\mathbf{B}^T)^{-1} - E\{\mathbf{f}(\mathbf{y})\mathbf{x}^T\} \quad (3.66)$$

3 Independent Component Analysis



(a)



(b)

Abbildung 3.10: Beispiele für den Zusammenhang zwischen Verteilungsdichten und den Randverteilungsdichten (a) Separierbare Verteilungsdichte und (b) nicht separierbare Verteilungsdichte.

3.3 ICA basierend auf der Kullback-Leibler-Divergenz

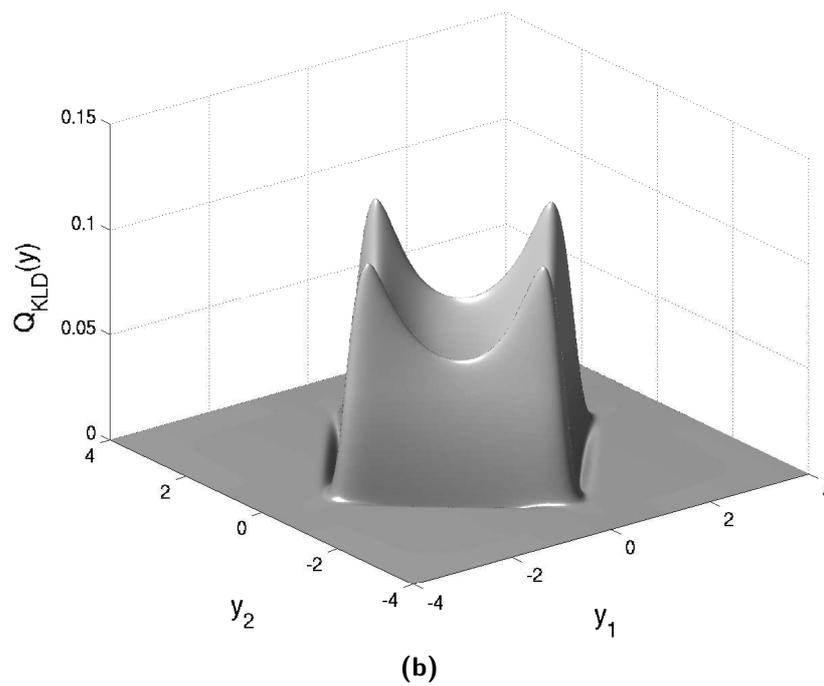
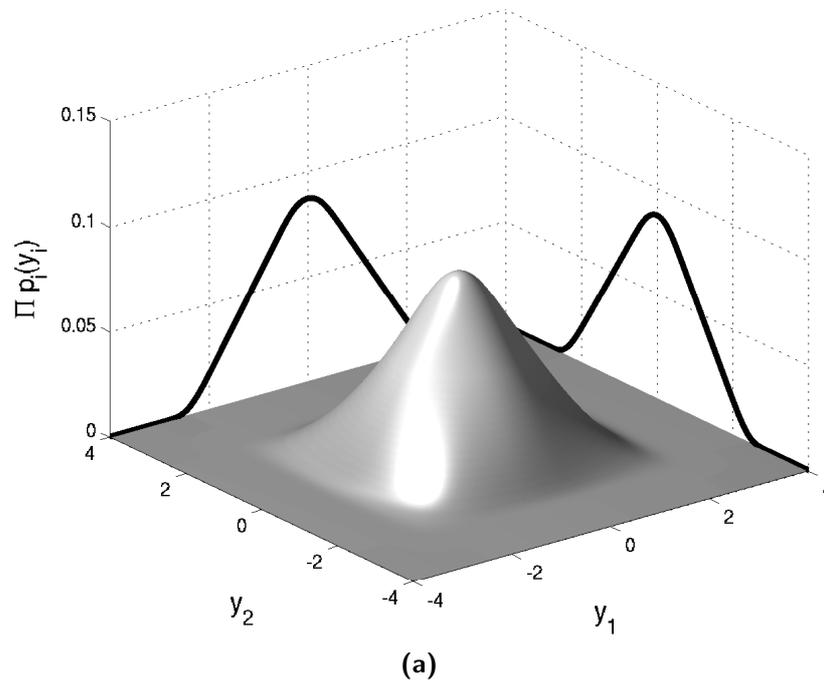


Abbildung 3.11: Zur Berechnung der KLD. (a) Verteilungsdichte, die sich aus der Multiplikation der Randverteilungen aus Abbildung 3.10 (b) ergibt. (b) Der Quotient aus Gleichung (3.57) gebildet zwischen der Verteilungsdichte aus Abbildung 3.11 (a) und 3.10 (b)

3 Independent Component Analysis

mit Hilfe der Matrixalgebra berechnet werden. Eine besondere Bedeutung hat in diesem Zusammenhang die *nichtlineare Aktivierungsfunktion*

$$\mathbf{f}(\mathbf{y}) = [f_1(y_1), f_2(y_2), \dots, f_N(y_N)]^T \quad (3.67)$$

mit

$$f_i(y_i) = -\frac{d \log p_i(y_i)}{dy_i} = -\frac{1}{p_i(y_i)} \cdot \frac{dp_i(y_i)}{dy_i} = -\frac{p_i'(y_i)}{p_i(y_i)}. \quad (3.68)$$

Dabei sind $p_i(y_i)$ die geschätzten Verteilungsdichten der einzelnen Komponenten. Auf die Möglichkeiten zur Schätzung dieser Verteilungsdichten wird im nächsten Kapitel genauer eingegangen.

Der Algorithmus aus Gleichung (3.66) wurde zuerst in [9] von Bell und Sejnowski vorgestellt. Der entscheidende Vorteil dieser Methode gegenüber FastICA ist die gleichzeitige Bestimmung aller Komponenten. Weiterhin ist sie wesentlich robuster als FastICA, da keine unsichere Schätzung der Kurtosis notwendig ist. Robustheit ergibt sich auch aus der Tatsache, dass die einzelnen Verteilungsdichten $p_i(y_i)$ nicht notwendigerweise exakt bekannt sein müssen. Kleine Fehler bei der Schätzung von $p_i(y_i)$ haben kaum einen Einfluß auf das Gesamtergebnis [4, 15, 13, 14, 31].

Natürlicher Gradient Der Algorithmus von Bell und Sejnowski funktioniert in der Praxis gut, er bedarf aber einer großen Anzahl an Iterationen und konvergiert deswegen unter Umständen langsam. Die Konvergenzgeschwindigkeit kann jedoch gesteigert werden, indem man weitere Eigenschaften, insbesondere die Nichtsingularität der Entmischungsmatrizen, ausnutzt. So bildet die Menge der nichtsingulären Matrizen eine multiplikative Gruppe. Weiterhin ist es eine Mannigfaltigkeit und eine Lie-Gruppe. Das bedeutet vor allem, dass der Raum der nichtsingulären Matrizen eine nicht euklidische Metrik besitzt. Deshalb ist der Gradient aus der Gleichung (3.64) nicht optimal.

Diese Tatsache wurde in [5, 3, 4] genutzt, um eine Riemann'sche Metrik für den Raum der nichtsingulären Matrizen einzuführen. Weiterhin wurde gezeigt, dass in diesem Raum der optimale Gradient der so genannte *natürliche Gradient*

$$\Delta_{\text{NG}} \mathbf{B} = -\frac{\partial \mathcal{R}(\mathbf{B}, \mathbf{y})}{\partial \mathbf{B}} \mathbf{B}^T \mathbf{B} \quad (3.69)$$

ist. Wie in [5] deutlich wurde, kann $\Delta_{\text{NG}} \mathbf{B}$ berechnet werden, indem der euklidische Gradient von rechts mit $\mathbf{B}^T \mathbf{B}$ multipliziert wird. Der Gradient aus der Gleichung (3.66) wird also zu

$$\Delta_{\text{NG}} \mathbf{B} = -\frac{\partial \mathcal{R}(\mathbf{B}, \mathbf{y})}{\partial \mathbf{B}} \mathbf{B}^T \mathbf{B} = [\mathbf{I} - E\{\mathbf{f}(\mathbf{y})\mathbf{y}^T\}] \mathbf{B}. \quad (3.70)$$

3.3 ICA basierend auf der Kullback-Leibler-Divergenz

Für den komplexen Fall muss die Transponierte durch die hermitesch Transponierte ersetzt werden:

$$\Delta_{\text{NG}} \mathbf{B} = [\mathbf{I} - E\{\mathbf{f}(\mathbf{y})\mathbf{y}^H\}] \mathbf{B}. \quad (3.71)$$

Der natürliche Gradient wurde parallel in [15] als sogenannter relativer Gradient eingeführt. Es existieren weitere Arbeiten, in denen der natürliche Gradient untersucht wurde [6, 20, 7, 2].

Ein weiterer positiver Aspekt der Formulierung (3.71) gegenüber (3.66) besteht in der vorteilhafteren Berechnungsmöglichkeit. So kann insbesondere die in (3.66) notwendige Matrixinvertierung vermieden werden, die unter Umständen numerisch problematisch sein kann.

Beispiel

Ein anschauliches Beispiel für den Unterschied zwischen dem euklidischen und natürlichen Gradienten ist in Abbildung 3.12 gegeben. Hier werden die kürzesten Wege auf einer Fläche verglichen. Ist die Fläche eine Kugeloberfläche, so ist der in rot dargestellte und mit Hilfe des natürlichen Gradienten berechnete Weg die kürzeste Verbindung zwischen zwei Punkten. Nach einer Abbildung auf eine euklidische Fläche ist hingegen der in blau dargestellte, nach dem euklidischen Gradienten berechnete Weg der kürzeste. Je nach Lage der Punkte können sich die Wege erheblich unterscheiden.

3.3.3 Aktivierungsfunktionen

Die Aktivierungsfunktionen aus Gleichung (3.66) beziehungsweise (3.67) hängen direkt von den Verteilungsdichten der einzelnen Signale ab:

$$f_i(y_i) = -\frac{p'_i(y_i)}{p_i(y_i)}. \quad (3.72)$$

Wie bereits erwähnt, haben kleine Fehler bei der Schätzung von $p_i(y_i)$ einen geringen Einfluss auf das Gesamtergebnis [4, 15, 13, 14, 31]. Trotzdem sollten die Verteilungsdichten nicht zu „falsch“ sein, da dann unter Umständen falsche Lösungen möglich sind [14].

Die Verteilungsdichten können wahlweise parametrisch oder nicht-parametrisch sein. Eine nicht-parametrische Verteilungsdichte kann zwar genauer sein, aber es

3 Independent Component Analysis

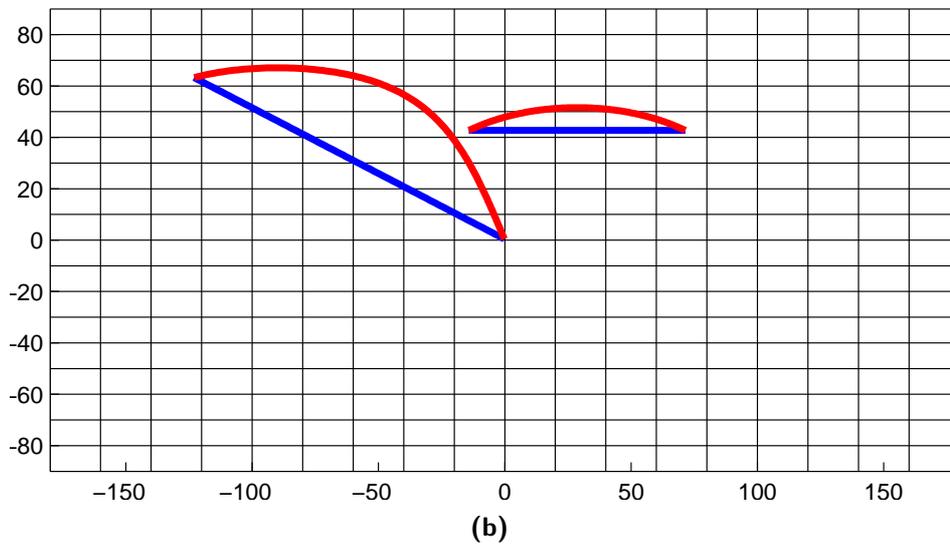
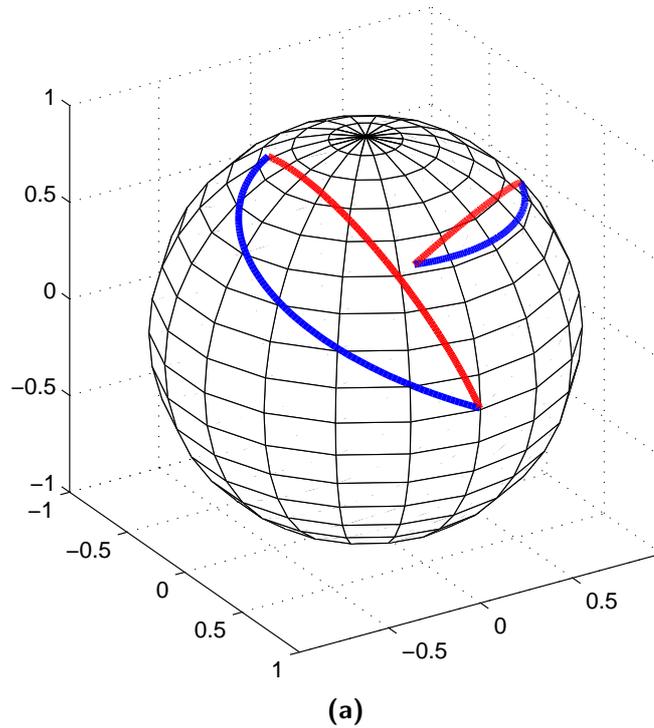


Abbildung 3.12: Beispiel für eine nichteuklidische Metrik. (a) Die kürzesten Wege zwischen zwei Punkten auf einer Kugeloberfläche sind in rot dargestellt. (b) Die kürzesten Verbindungen zwischen den gleichen Punkten bei einer euklidischen Fläche sind in blau markiert.

3.3 ICA basierend auf der Kullback-Leibler-Divergenz

ergeben sich Probleme bei zu wenigen Datenpunkten. Weiterhin muss die Ableitung der Verteilungsdichte sinnvoll definiert sein, so dass histogramm-basierte Methoden ungeeignet sind. Eine Schätzung über Parzen-Fenster ist möglich, aber auch hier bleibt die Frage der Fensterbreite bei wenigen Datenpunkten [66]. Weiterhin ist unter Umständen ein hoher Rechenaufwand zur Bestimmung von $f_i(y_i)$ nötig.

Die in dieser Arbeit schwerpunktmäßig betrachteten Sprachsignale können gut mittels parametrischer Verteilungsdichten beschrieben werden [26, 79]. Auch für die im nächsten Kapitel beschriebene Zeit-Frequenz-Darstellung lässt sich eine gute Beschreibung finden. Für diese Verteilungen ist auch eine Ableitung oftmals einfach berechenbar, und der Rechenaufwand für $f_i(y_i)$ ist deutlich kleiner. An dieser Stelle werden daher nur parametrische Verteilungsdichten berücksichtigt.

Laplaceverteilung

Für Sprachsignale kann eine Laplaceverteilung angenommen werden [26, 79]:

$$p_i(y_i) = \frac{1}{2\sigma} \exp\left(-\frac{|y_i|}{\sigma}\right). \quad (3.73)$$

Mit

$$p'_i(y_i) = -\frac{1}{2\sigma} \exp\left(-\frac{|y_i|}{\sigma}\right) \cdot \frac{\operatorname{sgn}(y_i)}{\sigma} \quad (3.74)$$

lautet die Aktivierungsfunktion:

$$f_i(y_i) = \frac{\operatorname{sgn}(y_i)}{\sigma}. \quad (3.75)$$

Zur Ableitung der Betragsfunktion wurde hier die sogenannte *schwache Ableitung für Distributionen* verwendet.

Kosinus Hyperbolicus

Eine ebenfalls häufig [29, 18] benutzte Aktivierungsfunktion ist der Tangens Hyperbolicus, der sich aus einer Unimodalen bzw. der Kosinus Hyperbolicus Verteilung ergibt.

Die Laplace- und Kosinus-Hyperbolicus-Verteilung sowie die dazugehörigen Aktivierungsfunktionen sind in Abbildung 3.13 dargestellt. Weitere Verteilungsdichten sowie die dazugehörigen Aktivierungsfunktionen sind in der Tabelle 3.1 zusammengefasst.

3 Independent Component Analysis

Tabelle 3.1: Beispiele für diverse Verteilungsdichten $p_y(y)$ und die dazugehörigen Aktivierungsfunktionen $f(y)$.

Name	Verteilungsdichte $p_y(y)$	Aktivierungsfunktion $f(y)$
Gauß	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{ y ^2}{2\sigma^2}\right)$	$\frac{y}{\sigma^2}$
Laplace	$\frac{1}{2\sigma} \exp\left(-\frac{ y }{\sigma}\right)$	$\frac{\text{sgn}(y)}{\sigma}$
Cauchy	$\frac{1}{\pi\sigma} \cdot \frac{1}{1 + \left(\frac{y}{\sigma}\right)^2}$	$\frac{2y}{\sigma^2 + y^2}$
Kosinus hyperbolicus	$\frac{1}{\pi \cosh\left(\frac{y}{\sigma^2}\right)}$	$\tanh\left(\frac{y}{\sigma^2}\right)$
Unimodal	$\frac{\exp(-2\sigma^{-2}y)}{(1 + \exp(-2\sigma^{-2}y))^2}$	$\tanh\left(\frac{y}{\sigma^2}\right)$
Dreieck	$\frac{1}{\sigma \left(1 - \frac{ y }{\sigma}\right)}$	$\frac{\text{sgn}(y)}{\sigma \left(1 - \frac{ y }{\sigma}\right)}$
Generalisierter Gauß	$\frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-\frac{ y }{\sigma}\right)^\beta$	$\frac{ y ^{\beta-1}}{\sigma^\beta} \text{sgn}(y)$

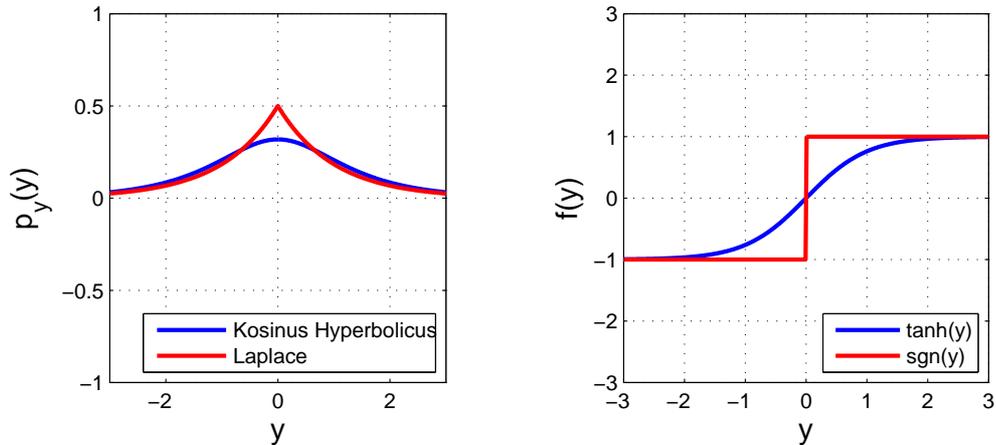


Abbildung 3.13: Zwei typische Verteilungsdichten und die dazugehörigen nicht-linearen Aktivierungsfunktionen.

Generalisierte Gaußdichte

Mit Hilfe der generalisierten Gaußdichte (*generalized gaussian distribution*, GGD) kann eine ganze Reihe von Verteilungsdichten beschrieben werden. Die mittelwertfreie generalisierte Gaußdichte wird durch die Standardabweichung $\sigma > 0$ und den Form-Parameter $\beta > 0$ bestimmt [60]:

$$p_i(y_i) = \frac{1}{2\Gamma(1 + \frac{1}{\beta})A(\beta, \sigma)} \exp\left(-\left|\frac{y_i}{A(\beta, \sigma)}\right|^\beta\right). \quad (3.76)$$

Hierbei ist $\Gamma(\cdot)$ die sogenannte Gamma-Funktion

$$\Gamma(y) = \int_0^\infty x^{y-1} e^{-x} dx \quad (3.77)$$

und $A(\beta, \sigma)$ ein Skalierungsfaktor, so dass die Bedeutung von σ^2 als Varianz mit $\text{Var}(y) = \sigma^2$ erhalten bleibt:

$$A(\beta, \sigma) = \left(\frac{\sigma^2 \Gamma(\frac{1}{\beta})}{\Gamma(\frac{3}{\beta})}\right)^{\frac{1}{2}}. \quad (3.78)$$

Mit dem Parameter β kann das Aussehen der Verteilung in einem sehr großen Bereich variiert werden. Für $\beta = 2$ reduziert sich die GGD zu der normalen Gaußdichte. Für $\beta = 1$ wird aus der GDD die Laplaceverteilung wie in Gleichung (3.73) definiert. Allgemein gilt, dass für $\beta < 2$ die Verteilungsdichte um den

3 Independent Component Analysis

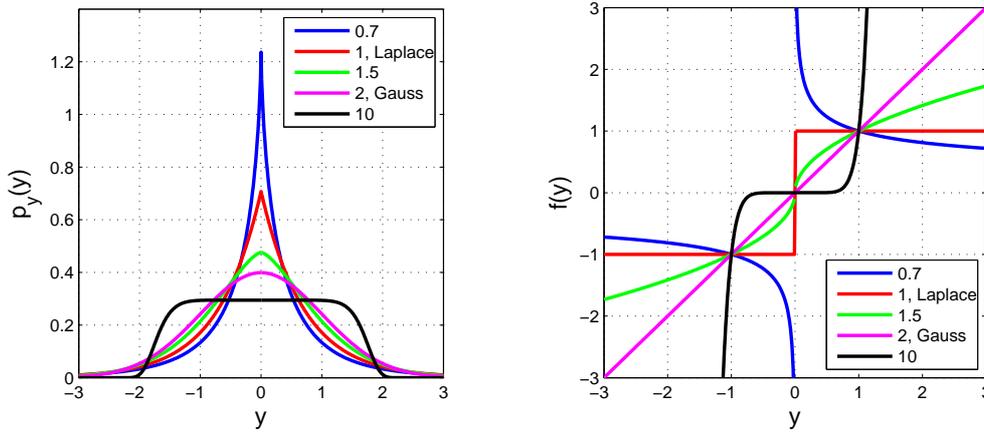


Abbildung 3.14: Die generalisierte Gaußdichte für verschiedene Parameter und die dazugehörigen Aktivierungsfunktionen.

Nullpunkt immer spitzer und an den Seiten angehoben wird. Für $\beta > 2$ werden die Seiten immer mehr abgesenkt und die Mitte immer flacher. Für den Grenzfall $\beta = \infty$ wird die GGD zu einer Gleichverteilung. In Abbildung 3.14 ist die GGD für verschiedene β dargestellt.

In diesem Zusammenhang kann auch die *generalisierte Varianz* (Dispersion) α^β als

$$\alpha^\beta = E\{|y|^\beta\} \quad (3.79)$$

definiert werden. Auch hierbei gilt, dass für die Normalverteilung mit $\beta = 2$ die Dispersion α^2 wiederum zur Varianz σ^2 wird.

Für den Fall, dass die Varianz nicht entscheidend ist, wird in der Literatur oft die folgende Variante der GGD verwendet [16, 18, 82]:

$$p_i(y_i) = \frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})} \exp\left(-\left|\frac{y_i}{\alpha}\right|^\beta\right). \quad (3.80)$$

Für beide Definitionen lautet die Aktivierungsfunktion:

$$f_i(y_i) = \frac{|y_i|^{\beta-1}}{\sigma^\beta} \text{sgn}(y_i) \quad (3.81)$$

und mit $\text{sgn}(y_i) = (y_i/|y_i|)$ ergibt sich die folgende Variante:

$$f_i(y_i) = \frac{1}{\sigma^\beta} \frac{y_i}{|y_i|^{2-\beta}}. \quad (3.82)$$

3.3 ICA basierend auf der Kullback-Leibler-Divergenz

Jede Komponente $y_i(n)$ kann dabei durchaus unterschiedliche β_i haben. Sind die β_i bekannt, so lassen sich alle notwendigen Parameter für (3.82) relativ schnell und einfach bestimmen. Damit kann das Gradientenverfahren nach Gleichung (3.70) problemlos durchgeführt werden. Bei einem vollständig blinden Verfahren sind die β_i jedoch nicht a priori bekannt und müssen geschätzt werden.

Sind die Signale bereits getrennt, ist eine Schätzung der einzelnen β_i kein Problem. Bevor das Gradientenverfahren jedoch konvergiert ist, ist eine Schätzung von β_i der i -ten getrennten Komponente nicht möglich. Da aber eine exakte Schätzung auch nicht nötig ist, ist es ausreichend, die Parameter aus den aktuellen, in der l -ten Iteration vorliegenden Signalen zu schätzen. Das führt zu folgendem alternierendem Verfahren:

1. Berechne mit Hilfe der vorliegenden Trennmatrix \mathbf{B}^l die getrennten Signale $\mathbf{y}(n)$. Schätze β_i für alle möglicherweise nicht gut getrennten Komponenten.
2. Führe einen Iterationsschritt nach Gleichung (3.70) aus. Sofern noch keine Konvergenz vorliegt, gehe zu 1.

Nachdem der Algorithmus konvergiert ist, liegt eine Trennmatrix \mathbf{B} samt der getrennten Signale $\mathbf{y}(n)$, sowie eine Schätzung der β_i für alle Komponenten vor.

Solche adaptiven Algorithmen wurden bereits in [16, 40] vorgestellt. Weiterhin wurde dort gezeigt, dass auf diese Art und Weise auch Signale mit sehr unterschiedlichen β_i getrennt werden können. Insgesamt sind derartige adaptive Algorithmen sehr robust. Nachteilig ist der große Rechenaufwand für die Schätzung der β_i in jedem Iterationsschritt. Das Verfahren kann beschleunigt werden, indem zum Beispiel auf die Schätzung der Parameter in jedem einzelnen Schritt verzichtet wird und die Werte aus der vorherigen Iteration benutzt werden.

Schätzer für die Parameter der GGD

Für das adaptive Verfahren müssen die Parameter der GGD in jedem Iterationsschritt bestimmt werden. Die Schätzung dieser Parameter aus einer endlichen Stichprobe ist unterschiedlich aufwändig.

Bei Mittelwertfreiheit ist die Varianz σ^2 das zweite Moment

$$\sigma^2 = E\{y^2\} = \frac{1}{N} \sum_{i=1}^N y^2. \quad (3.83)$$

3 Independent Component Analysis

Für die Dispersion gilt bei bekannten β :

$$\alpha^\beta = E\{|y|^\beta\} = \frac{1}{N} \sum_{i=1}^N |y|^\beta. \quad (3.84)$$

Die Schätzung des Form-Parameters β ist hingegen deutlich schwieriger. Es wurden mehrere Methoden vorgestellt [41, 44, 78]. Ein Vergleich [41] zeigt, dass der *moment matching estimator (MME)* [44] eine gute Schätzung liefert und sich besonders schnell berechnen lässt. Alternativ kann auch der *entropy matching estimator (EME)* benutzt werden. Die Berechnung ist ähnlich schnell, jedoch ist der Schätzer, besonders bei kleinen Stichproben, deutlich schlechter. Der *maximum likelihood estimator (MLE)* ist am genauesten, jedoch ist der Rechenaufwand extrem groß, da keine geschlossene Form existiert und somit eine numerische Optimierung notwendig ist.

Die Idee des MME ist es, den Form-Parameter so zu wählen, dass alle Momente des Datensatzes gleich den Momenten der modellierten GGD sind. Da die GGD symmetrisch ist, können dazu nur die geraden Momente benutzt werden. In [78] wurde dazu vorgeschlagen, lediglich zwei Momente vom mindestens viertem Grade anzupassen. Da die Bestimmung solch hoher Momente aber sehr unsicher ist, ist die Schätzung von β ebenfalls unsicher. Eine bessere Methode ist es, die absoluten Momente zu benutzen, da dann auch ungerade Momente und Momente kleiner Ordnung benutzt werden können.

Das r -te absolute Moment ist definiert als

$$E\{|y|^r\} = \int_{-\infty}^{\infty} |y|^r p_y(y) dy. \quad (3.85)$$

Für die GGD ergibt sich damit

$$m_r = E\{|y|^r\} = A^r(\beta, \sigma) \frac{\Gamma\left(\frac{r+1}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)}. \quad (3.86)$$

In [44] wurde das Verhältnis zwischen dem Quadrat des ersten Moments und dem zweiten absoluten Moment

$$\mathcal{M}(\beta) = \frac{m_1^2}{m_2} = \frac{\Gamma^2\left(\frac{2}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)\Gamma\left(\frac{3}{\beta}\right)} \quad (3.87)$$

benutzt um β zu schätzen:

$$\beta = \mathcal{M}^{-1}\left(\frac{E\{|y|\}^2}{E\{|y|^2\}}\right). \quad (3.88)$$

3.3 ICA basierend auf der Kullback-Leibler-Divergenz

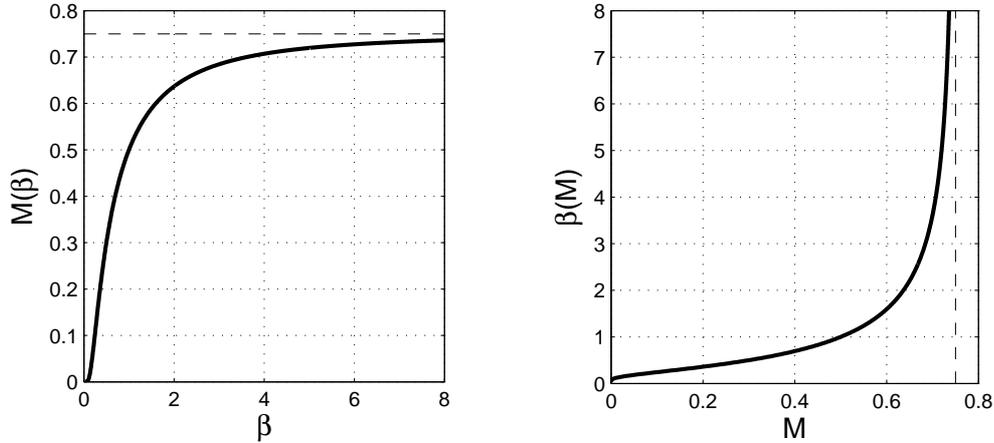


Abbildung 3.15: Schätzung des Form-Parameters β . Auf der linken Seite $M(\beta)$ und rechts die approximierte *generalized Gaussian ratio function*.

$\mathcal{M}^{-1}(\cdot)$ ist auch als *generalized Gaussian ratio function* bekannt.

Alternativ wurden in [17] Varianz und das vierte Moment benutzt:

$$\mathcal{K}(\beta) = \frac{m_2^2}{m_4} = \frac{\Gamma^2\left(\frac{3}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)\Gamma\left(\frac{5}{\beta}\right)}. \quad (3.89)$$

Auch hier kann β durch Invertieren berechnet werden:

$$\beta = \mathcal{K}^{-1}\left(\frac{E\{|y|^2\}^2}{E\{|y|^4\}}\right). \quad (3.90)$$

Die Bestimmung des vierten Moments ist bei kleinen Datensätzen ungenau, daher sollte an dieser Stelle $\mathcal{M}(\cdot)$ bevorzugt werden.

In Abbildung 3.15 ist $\mathcal{M}(\cdot)$ dargestellt. Die Funktion ist streng monoton steigend. Das Inverse existiert zwar, aber es gibt dafür keine bekannte geschlossene Form. Eine praktische Lösung für die Berechnung ist zum Beispiel eine interpolierte Lookup-Tabelle. Laut [41] ist an dieser Stelle eine Genauigkeit von 10^{-3} ausreichend. Alternativ kann auch die in [60] vorgeschlagene Approximation durch abschnittsweise definierte Funktionen benutzt werden. Diese Approximation, in Abbildung 3.15 dargestellt, hat die geforderte Genauigkeit.

Die MME-Methode erlaubt eine gute Schätzung des Form-Parameters in einem Bereich zwischen ungefähr 0.05 und 8 und ist somit mehr als ausreichend für typische Sprachsignale [12].

3.3.4 Komplexwertige Signale

Die ICA für komplexwertige Signale ist grundsätzlich vergleichbar mit der für reellwertige Signale. In Gleichung (3.71) wurde die allgemeine Form für diesen Fall bereits definiert. Allerdings ist zu beachten, dass die Verteilungsdichten $p_i(y_i)$ nun Funktionen komplexwertiger Variablen sind, also besonders betrachtet werden müssen. Dies gilt insbesondere für die Aktivierungsfunktionen $f_i(y_i)$, da diese die Ableitung $p'_i(y_i)$ beinhalten, welche nur existiert, wenn $p_i(y_i)$ holomorph ist. Eine einfache Zahlenbereichserweiterung der bisher besprochenen, reellen Verteilungsdichten auf die komplexe Ebene führt dabei nicht zu holomorphen Funktionen, so dass hier eine andere Vorgehensweise nötig ist. Im Folgenden werden zwei Ansätze besprochen: Einerseits die Annahme der Unabhängigkeit des Real- und Imaginärteils und andererseits die zirkulärsymmetrischen Verteilungsdichten.

Komplexwertige Verteilungsdichten

Eine komplexwertige Funktion kann in den Real- und Imaginärteil

$$f(z) = u(x, y) + jv(x, y) \quad (3.91)$$

mit $z = x + jy$ aufgespalten werden. Die Ableitung existiert genau dann, wenn die Cauchy-Riemannschen Gleichungen

$$\frac{du}{dx} = \frac{dv}{dy}, \quad \frac{dv}{dx} = -\frac{du}{dy} \quad (3.92)$$

erfüllt sind. Diese Forderungen sind wesentlich stärker als die für die reellwertige Differenzierbarkeit. Insbesondere gilt für alle Funktionen $f : \mathbb{C} \rightarrow \mathbb{R}$, also auch für komplexe Verteilungsdichten

$$v(x, y) = 0, \quad (3.93)$$

so dass diese nicht differenzierbar sind. Basierend auf einer weniger strengen Definition der Differenzierbarkeit [81, 43] wurden in [10] und [77] partielle Ableitungen benutzt. Mit

$$x = \frac{1}{2}(z + z^*), \quad y = -j\frac{1}{2}(z - z^*) \quad (3.94)$$

und der reellen Differenzierbarkeit von $u(x, y)$ und $v(x, y)$ nach beiden Variablen kann mit einfacher Substitution eine Differenzierbarkeit nach z und z^* berechnet

3.3 ICA basierend auf der Kullback-Leibler-Divergenz

werden. Mit $g(z, z^*) = f(x, y)$ ergibt sich:

$$\frac{dg}{dz} = \frac{1}{2} \left(\frac{df}{dx} - j \frac{df}{dy} \right), \quad \frac{dg}{dz^*} = \frac{1}{2} \left(\frac{df}{dx} + j \frac{df}{dy} \right). \quad (3.95)$$

Hierbei wird jeweils z beziehungsweise z^* als konstant angenommen. Damit kann nun zum Beispiel

$$\frac{df}{dz} = \frac{1}{2} \left(\frac{dg}{dz} + j \frac{dg}{dz^*} \right) \quad (3.96)$$

als Ableitung definiert werden. Im Falle einer holomorphen Funktion ist diese Ableitung gleich der komplexen Ableitung.

Komplexwertige Zufallsvariablen

Das einfachste Modell für komplexwertige Zufallsvariablen besteht aus unabhängigen Real- und Imaginärteilen, welche wiederum eine reelle Verteilungsdichte haben. In der Abbildung 3.16 (a) ist ein einfaches Modell mit jeweils einer Laplaceverteilung für beide Komponenten dargestellt:

$$p_y(y) = \frac{1}{4\sigma^2} \exp\left(-\frac{|\mathcal{R}(y)|}{\sigma}\right) \cdot \exp\left(-\frac{|\mathcal{I}(y)|}{\sigma}\right). \quad (3.97)$$

Die Aktivierungsfunktion hat dann die Form

$$f(y) = \text{sgn}(\mathcal{R}(y)) \cdot \text{sgn}(\mathcal{I}(y)). \quad (3.98)$$

Dieses einfache Modell ist nicht geeignet, um eine einzelne Frequenz aus einer Zeit-Frequenz Darstellung, wie im nächsten Kapitel beschrieben, zu modellieren. Hier sind der Real- und Imaginärteil nicht unabhängig. Vielmehr repräsentiert die Phase lediglich eine Zeitverschiebung, so dass jede Phase gleich wahrscheinlich ist. Damit werden die Real- und Imaginärteile voneinander abhängig. Es muss lediglich eine Modellierung für den Betrag gefunden werden. Diese Überlegungen führen zu den sogenannten *zirkulär symmetrischen Verteilungen* (spherically invariant distributions) [11, 12, 42]. Bei diesen Verteilungen wird, unabhängig von der Phase, der Funktionswert nur über den Abstand zum Ursprung definiert.

Die Laplaceverteilung aus Gleichung (3.97) wird bei dieser Modellierung zu

$$p_y(y) = \frac{1}{2\sigma} \exp\left(-\frac{|y|}{\sigma}\right). \quad (3.99)$$

3 Independent Component Analysis

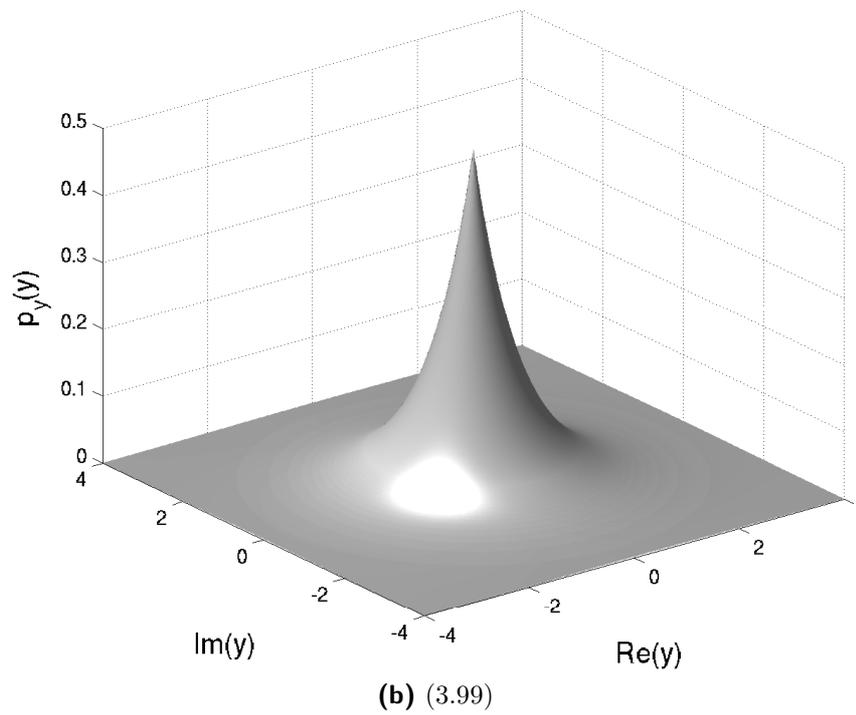
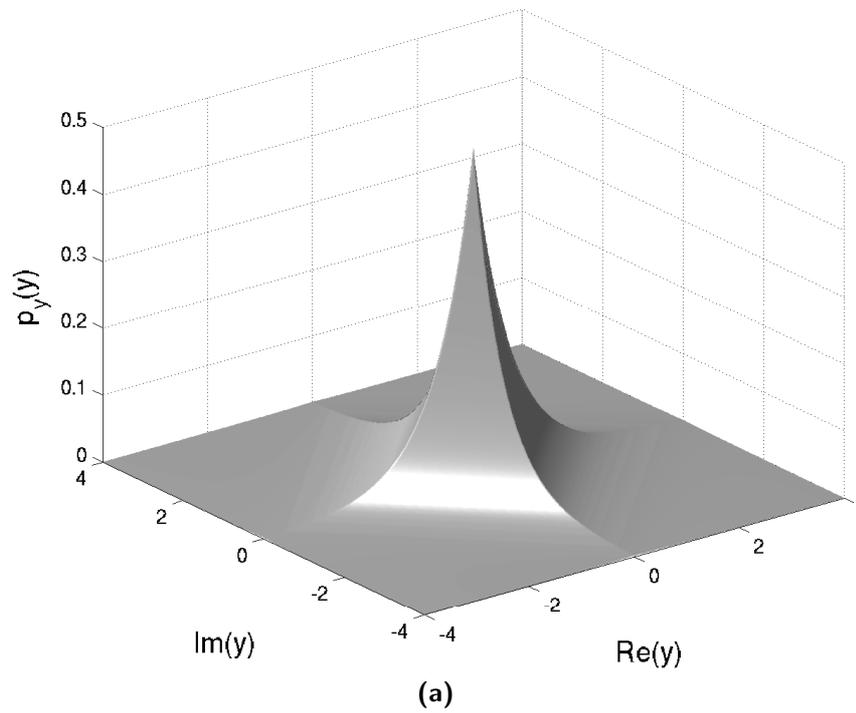


Abbildung 3.16: Verteilungsdichten komplexwertiger Zufallsvariablen. (a) Verteilungsdichte nach Gleichung (3.97). (b) Verteilungsdichte nach Gleichung (3.99)

Diese Formulierung ist syntaktisch gleich dem reellwertigen Modell, jedoch ist hier y komplexwertig. Ein Beispiel für diese rotationssymmetrische Verteilungsdichte ist in Abbildung 3.16 (b) gegeben.

Bei dieser Formulierung muss die Bestimmung der Ableitung für die Aktivierungsfunktion entsprechend angepasst werden. Statt einer Berechnung nach (3.96) sollte die Nichtlinearität nur auf den Betrag angewendet werden [42, 11]. Für die Laplaceverteilung gilt also

$$f(y) = \operatorname{sgn}(y) = \frac{y}{|y|}. \quad (3.100)$$

Für die GGD ergibt sich das gleiche Bild. Auch hier können sowohl die Definition der Verteilungsdichte aus Gleichung (3.80) wie auch der nichtlinearen Aktivierungsfunktion aus Gleichung (3.82) übernommen werden, sofern die Definition des komplexen Vorzeichens aus (3.100) übernommen wird.

Schätzer für die komplexe GGD Durch die konsistente Erweiterung der GGD auf die komplexe Ebene muss das Schätzverfahren für die Parameter der GGD nicht verändert werden. Insbesondere gilt aufgrund der zirkulären Symmetrie, dass die Parameter der GGD ausgehend von dem Ursprung in alle Richtungen gleich sind. Dies passt sehr gut zu dem Schätzer aus Gleichung (3.88), der über die absoluten Momente definiert ist und somit unverändert übernommen werden kann.

Beispiel

In Abbildung 3.17 ist ein Beispiel für eine zirkulär symmetrische GGD mit $\beta = 10$ dargestellt. Zusätzlich sind die reellen GGD entlang der Koordinatenachsen eingezeichnet, mit deren Hilfe die Parameter der GGD bestimmt werden können.

3.4 Zusammenfassung

In diesem Kapitel wurden Methoden vorgestellt, mit denen lineare instantane Mischungen von Signalen getrennt werden können. Sie eignen sich bei nicht-gaußschen Signalen und nutzen die Tatsache aus, dass Mischungen von Signalen eine Verteilungsdichte haben, die näher an der Gaußverteilung ist. Durch die

3 Independent Component Analysis

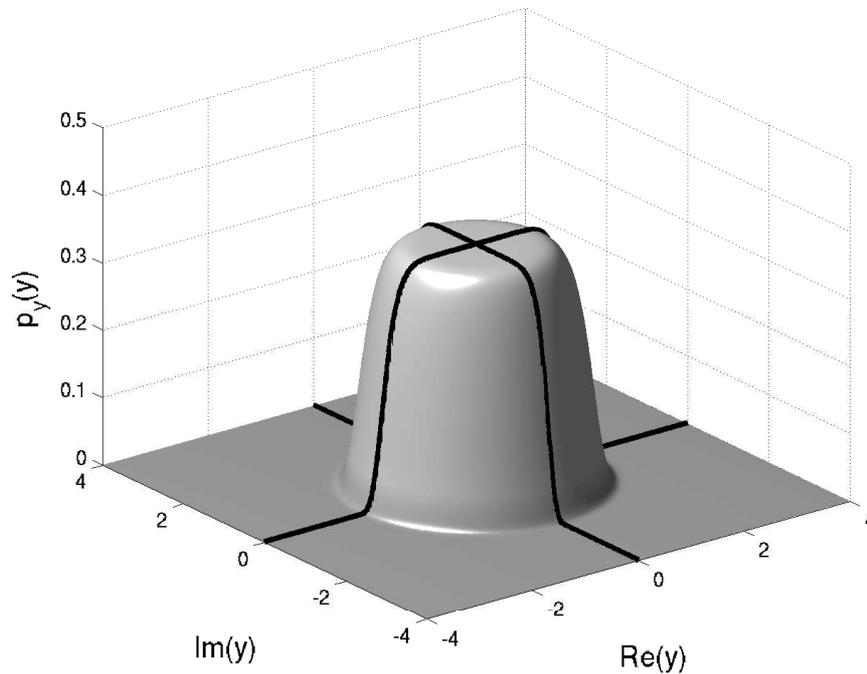


Abbildung 3.17: Zirkulär symmetrische GGD mit $\beta = 10$.

Messung der Gaußheit mit Hilfe der Kurtosis können einzelne Quellen identifiziert und mit Hilfe einer Projektion extrahiert werden. Die zweite vorgestellte Methode basiert auf der Tatsache, dass Verteilungsdichten für unabhängige Komponenten in einzelne Randverteilungsdichten separiert werden können. Dieser Ansatz führt zu dem Bell-Sejenowski-Algorithmus, der ebenfalls leicht für komplexwertige Signale erweitert werden kann. Für die bei dem Bell-Sejenowski-Algorithmus benötigten Aktivierungsfunktionen wurde ein adaptives Modell basierend auf der generalisierten Gaußdichte vorgestellt.

Kapitel 4

Konvolutive Mischungen

Im letzten Kapitel wurde eine Methode vorgestellt, mit deren Hilfe es möglich ist, instantane Mischungen zu trennen. Das vorgestellte Modell ist adaptiv und erlaubt die Trennung von Signalen, die eine nichtgaußsche Statistik besitzen, wie zum Beispiel Sprachsignale. Die Statistik der Signale kann mit Hilfe der GGD modelliert werden. Ebenfalls wurde dazu ein Schätzverfahren für komplexe Signale vorgestellt.

Handelt es sich jedoch um Aufnahmen in realen Räumen, also um eine akustische Mischung, ist dieses Modell nicht ausreichend. Die endliche Schallgeschwindigkeit hat in solch einer Umgebung zur Folge, dass die Signale verzögert bei den Empfängern ankommen. Weiterhin werden die Signale an den Wänden und weiteren möglichen Gegenständen im Raum reflektiert, so dass die Signale wiederholt und verzögert auf die Mikrofone auftreffen. Dieses Verhalten kann mit Hilfe einer Impulsantwort modelliert werden. Um einen typischen Raum zu modellieren, sind jedoch Impulsantworten mit einer Länge von mehreren Tausend Koeffizienten notwendig. Insgesamt wird das Mischungssystem nicht mehr mit Hilfe einer Matrix \mathbf{A} beschrieben, sondern mit einem System bestehend aus einer Folge von Matrizen $\mathbf{H}(n)$. Bei N Quellen und N Mikrofonen und einer Impulsantwortlänge L_H werden für die Modellierung also $N^2 L_H$ Koeffizienten benötigt.

Um eine Trennung solcherart gemischter Signale zu erreichen, ist eine Invertierung dieses Mischungssystems notwendig. Diese Invertierung bedarf eines Entmischungssystems gleicher Größenordnung. Das bedeutet insbesondere, dass für ein einfaches (2×2) -System bei einer Impulsantwortlänge von 2000 Koeffizienten bereits 8000 Werte geschätzt werden müssen. Es ist möglich, alle diese Koeffizienten direkt im Zeitbereich zu schätzen [25, 1], doch sind die Ergebnisse oft

4 Konvolutive Mischungen

nicht zufriedenstellend. Einerseits ist ein extremer Rechenaufwand nötig, und andererseits bleiben die Algorithmen in lokalen Minima gefangen.

Ein besserer Ansatz ist es, das Problem im Frequenzbereich zu formulieren. Auf diese Weise wird die Faltung mit der Impulsantwort in eine Multiplikation transformiert. Das erlaubt dann eine Trennung jeder Frequenz unabhängig von jeder anderen. Die dazu notwendigen Methoden wurden schon im vorherigen Kapitel vorgestellt. Dieser Ansatz ist jedoch mit einem nicht vernachlässigbaren Nachteil verbunden. Die ICA kann weder die Reihenfolge, noch die Energie der Signale bestimmen. Somit können die Signale in den einzelnen Frequenzbändern beliebig skaliert und permutiert werden. Ohne eine Korrektur der Permutationen ist das rekonstruierte Signal nicht getrennt und eine fehlende Korrektur der Skalierung führt zu einer gefilterten Version, die sich erheblich von dem Original unterscheiden kann. Neue Ansätze für die Lösung dieser Probleme werden in den Kapiteln 5 und 6 vorgestellt.

In den folgenden Abschnitten dieses Kapitels wird zuerst darauf eingegangen, wie eine Raumimpulsantwort modelliert werden kann und was das für das Mischungssystem bedeutet. Danach wird gezeigt, wie mit Hilfe der Kurzzeit-Fourier-Transformation der Übergang in den Zeit-Frequenzbereich gemacht werden kann und welche Nebenbedingungen eingehalten werden müssen. Schließlich werden noch Bewertungsfunktionen vorgestellt, mit denen die Qualität der Trennung gemessen werden kann.

4.1 Die Raumimpulsantwort

Die Raumimpulsantwort modelliert das akustische Verhalten eines Raumes. Grundsätzlich existieren zwei wichtige Modelle: Einerseits das Freifeldmodell aus Abbildung 4.2 (a) und andererseits die Modellierung von Räumen mit Reflexionen aus Abbildung 4.2 (b). Das Freifeldmodell ist dabei sehr einfach, da es durch einen einzigen Dirac-Impuls approximiert werden kann. Die Verschiebung des Impulses ist dabei direkt abhängig von der Entfernung der Quelle zum Mikrofon und somit der Laufzeit des Schalls.

Werden hingegen abgeschlossene Räume modelliert, müssen zusätzlich noch Reflexionen berücksichtigt werden wie in Abbildung 4.1 (b) dargestellt. Dominierend ist dabei nach wie vor der direkte Weg, der bei einer Impulsantwort, wie zum Beispiel in Abbildung 4.2 (a) dargestellt, als erstes und größtes Maximum deutlich sichtbar ist. Jedes reflektierte Signal führt zu einem weiteren Ausschlag in der Raumimpulsantwort. Eine Reflexion bedingt, abhängig von physikalischen

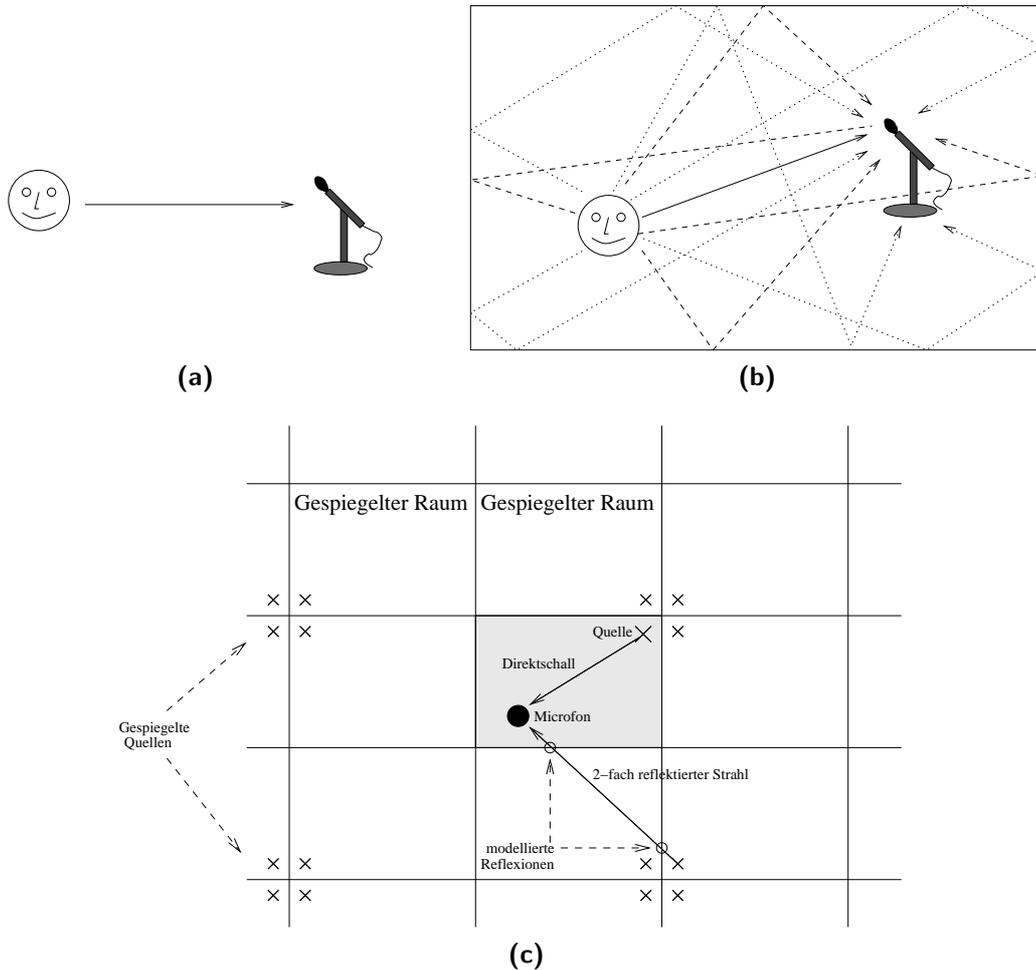
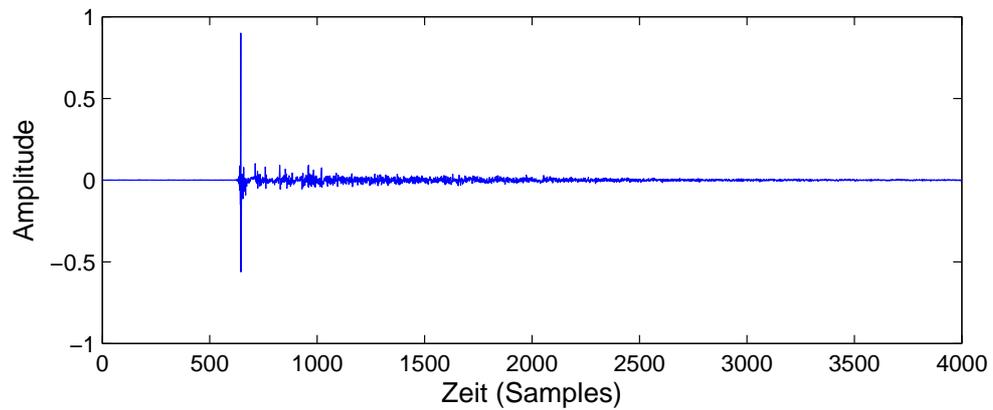


Abbildung 4.1: Unterschiedliche Modelle für Raumimpulsantworten. (a) Das Freifeldmodell mit Verzögerung aufgrund der Laufzeit. (b) Das Raummodell mit vielen zusätzlichen Wiederholungen aufgrund von Reflexionen. (c) Spiegelbildmethode zur Simulation von Raumimpulsantworten.

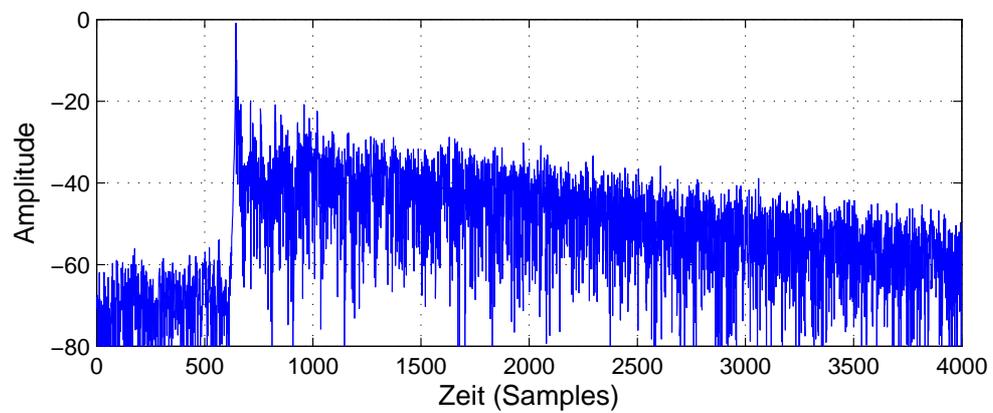
Eigenschaften des Materials, gleichzeitig auch einen Verlust an Energie. Dieser Verlust, der mit einem Faktor γ_r beschrieben werden kann, führt bei einer Mehrfachreflexion zu einer exponentiellen Abnahme der Energie. Diese Eigenschaft kann sehr gut in der Abbildung 4.2 (b) beobachtet werden, wo die Energie auf einer logarithmischen Skala nahezu linear abnimmt.

Ein einfaches Verfahren für die Modellierung einer Impulsantwort ist in Abbildung 4.1 (c) gegeben. Bei der sogenannten Spiegelbildmethode wird der Raum inklusive der Quelle unendlich oft in alle Richtungen gespiegelt. Mit Hilfe aller

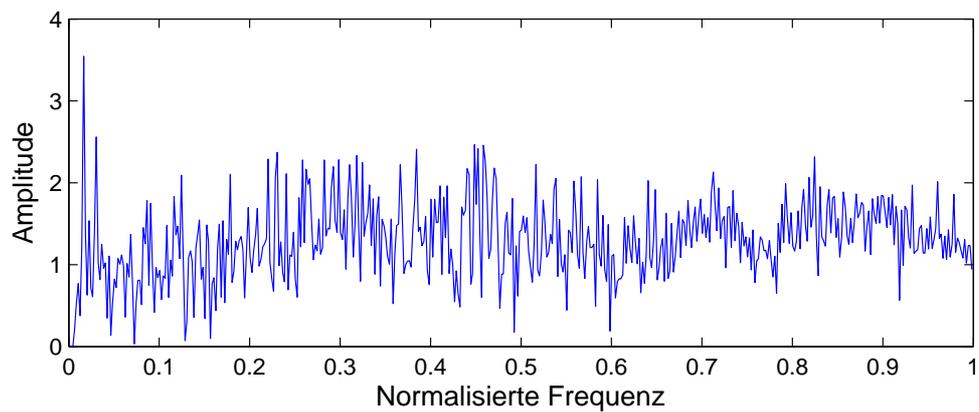
4 Konvolute Mischungen



(a)



(b)



(c)

Abbildung 4.2: Verschiedene Darstellungsformen einer Impulsantwort. (a) Eine Raumimpulsantwort im Zeitbereich. (b) Die Energie der einzelnen Samples einer Impulsantwort. (c) Betragsfrequenzgang.

Spiegelquellen wird die Impulsantwort zusammengesetzt. Aus dem Abstand der Spiegelquelle zum Mikrofon ergibt sich die Verzögerung und aus der Anzahl der Reflexionen die Energie [37].

Insgesamt kann eine Raumimpulsantwort in drei Bereiche gegliedert werden. Der erste Teil besteht aus einem verschobenen Dirac-Impuls, der den Direktschall repräsentiert. Es folgen die sogenannten frühen Reflexionen. Diese können oftmals direkt einer bestimmten Spiegelung zugeordnet werden. Mit fortlaufender Zeit steigt auch die Anzahl an reflektierten Signalen, welche sich immer mehr überlappen und nicht genau zugeordnet werden können. Dies ist der sogenannte diffuse Nachhall.

Eine Raumimpulsantwort kann, wie bereits gezeigt, in der Zeitdomäne als eine Folge von Koeffizienten beschrieben werden. Sie ist zwar nach den theoretischen Überlegungen unendlich lang, kann aber durch die ersten L_H Koeffizienten hinreichend gut beschrieben werden. Für reale Räume hängt L_H von der Abtastrate und der Nachhallzeit ab und beträgt oftmals mehrere Tausend. Die Wirkung einer Raumimpulsantwort $h(n)$ mit der Länge L_H auf ein Signal $s(n)$ ist durch die Faltung gegeben:

$$x(n) = h(n) * s(n) = \sum_{l=0}^{L_H-1} h(l)s(n-l). \quad (4.1)$$

Für die blinde Quellentrennung ist die Frequenzdarstellung von besonderem Interesse. Diese wird durch die zeitdiskrete Fourier-Transformation (DTFT) erreicht:

$$H(e^{j\omega}) = \text{DTFT}(h(n)). \quad (4.2)$$

Die Faltung wird im Frequenzbereich zu einer Multiplikation:

$$X(e^{j\omega}) = H(e^{j\omega}) \cdot S(e^{j\omega}). \quad (4.3)$$

Eine kleine Änderung der Position von Quelle oder Mikrofon relativ zueinander ist ausreichend, um die Impulsantwort entscheidend zu verändern. Sind mehrere Quellen und Sensoren vorhanden, muss jeder Übertragungsweg einzeln modelliert werden. Diese Impulsantworten $h_{ij}(n)$, welche die Übertragungsfunktion von der j -ten Quelle zu dem i -ten Mikrofon beschreiben, können dabei in einer Matrix

$$\mathbf{H}(n) = \begin{bmatrix} h_{11}(n) & h_{12}(n) & \dots & h_{1N}(n) \\ h_{21}(n) & h_{22}(n) & \dots & h_{2N}(n) \\ \dots & \dots & \dots & \dots \\ h_{N1}(n) & h_{N2}(n) & \dots & h_{N,N}(n) \end{bmatrix} \quad (4.4)$$

4 Konvolutive Mischungen

zusammengefasst werden. Das gesamte Mischungssystem lautet also

$$\mathbf{x}(n) = \mathbf{H}(n) * \mathbf{s}(n) = \sum_{l=0}^{L_H-1} \mathbf{H}(l) \mathbf{s}(n-l). \quad (4.5)$$

Auch hier ist eine Beschreibung im Frequenzbereich mit

$$\mathbf{X}(e^{j\omega}) = \mathbf{H}(e^{j\omega}) \cdot \mathbf{S}(e^{j\omega}) \quad (4.6)$$

möglich. Das inverse System

$$\mathbf{W}(e^{j\omega}) = \mathbf{H}^{-1}(e^{j\omega}) \quad (4.7)$$

ist das gesuchte Entmischungssystem. Da die Formulierung aus Gleichung (4.6) für jede Frequenz nur einen Wert liefert, ist die Bestimmung von $\mathbf{W}(e^{j\omega})$ mit den bisherigen Mitteln nicht möglich. Aus diesem Grund wird in der Literatur oft die Kurzzeit-Fouriertransformation benutzt [32, 61, 73], welche eine wesentlich einfachere Bestimmung des inversen Systems erlaubt.

4.2 Die Kurzzeit-Fouriertransformation

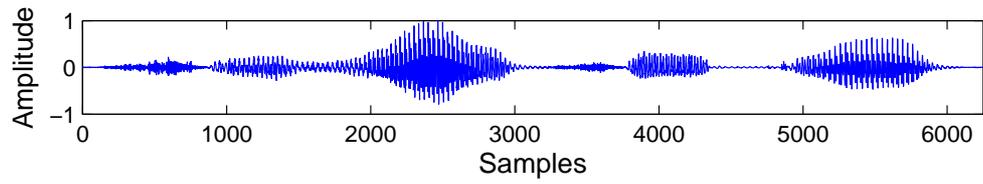
Mit Hilfe der *Kurzzeit-Fouriertransformation* (short time Fourier transform, STFT) kann der Verlauf der einzelnen Frequenzkomponenten eines Signals $x(n)$ über die Zeit untersucht werden. Weiterhin ermöglicht die STFT eine Verarbeitung im Frequenzbereich. Für die blinde Quellentrennung ist diese Darstellung von großer Bedeutung, da sie eine einfache Trennung in jedem Frequenzband erlaubt [32, 61, 73, 69, 33].

Für die Berechnung der STFT wird das Signal in kleine Abschnitte unterteilt, mit einer Fensterfunktion gewichtet und fouriertransformiert. Der Betrag der Kurzzeit-Fouriertransformierten ist das sogenannte *Spektrogramm*. Es erlaubt eine sehr anschauliche Interpretation der Frequenzverläufe. Im Folgenden werden die einzelnen Schritte erklärt und die Bedingungen für eine korrekte Rücktransformation definiert.

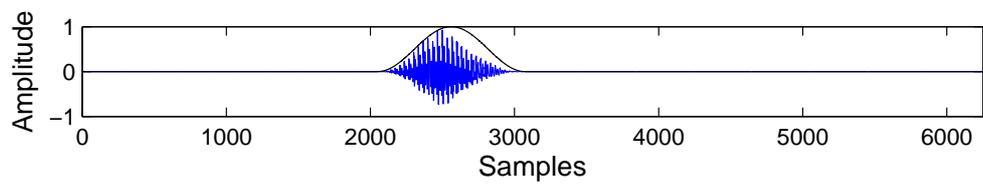
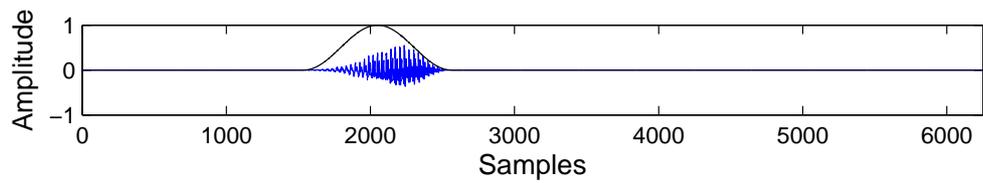
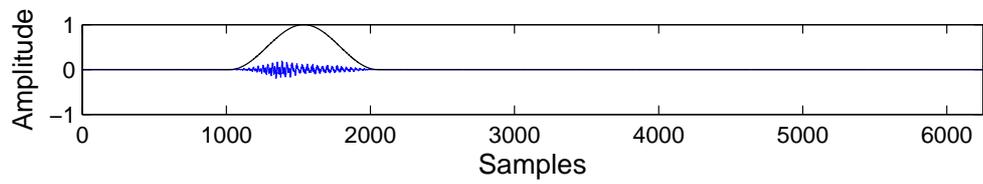
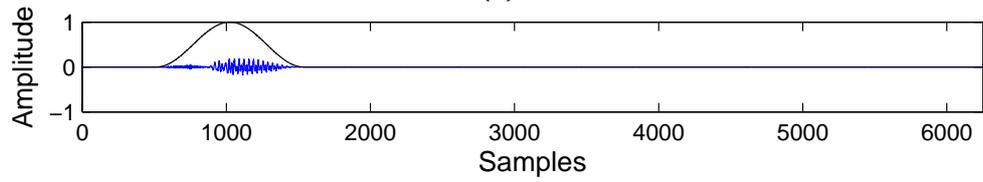
Fensterung

Der erste Schritt der STFT ist die Blockbildung und Fensterung, wie in Abbildung 4.3 dargestellt. Dazu wird ein Block der Länge L_γ aus dem Signal heraus-

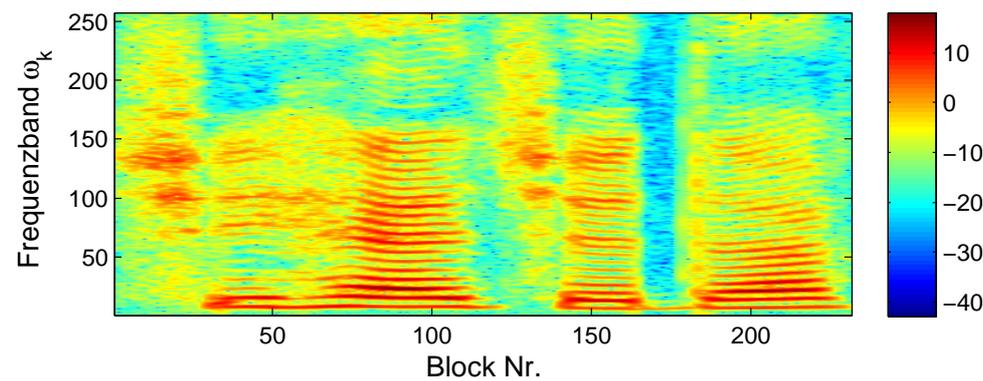
4.2 Die Kurzzeit-Fouriertransformation



(a)



(b)



(c)

Abbildung 4.3: Die Berechnung eines Spektrogramms. (a) Ein Sprachsignal. (b) Gefensterte Abschnitte. (c) Das fertige Spektrogramm.

4 Konvolutive Mischungen

geschnitten und mit einem Fenster $\gamma(n)$ gleicher Länge gewichtet:

$$x_b(n, \tau) = x(n + \tau L_V) \gamma(n), \quad 0 \leq n \leq L_\gamma - 1. \quad (4.8)$$

Dabei ist τ die Blocknummer und L_V der Vorschub. Die nacheinander folgenden Blöcke haben eine Überlappung von $L_\gamma - L_V$.

Frequenztransformation

Im nächsten Schritt werden die einzelnen Blöcke mit Hilfe der diskreten Fouriertransformation (DFT) in den Frequenzbereich überführt. Mit der Fouriermatrix \mathcal{F} gilt

$$X(\omega_k, \tau) = \mathcal{F} x_b(n, \tau). \quad (4.9)$$

Vor der DFT kann $x_b(n, \tau)$ um L_H Nullen erweitert werden (*zero padding*), um eine höhere Auflösung zu erreichen. Die Länge der DFT ist

$$L_{\text{FFT}} = L_\gamma + L_H. \quad (4.10)$$

Bei der Implementierung kann natürlich auch die schnellere FFT benutzt werden. Das Ergebnis ist eine zweidimensionale Zeit-Frequenz-Darstellung, die angibt, wie stark eine Frequenz ω_k im Block τ vorhanden ist. In der Abbildung 4.3 (c) ist das Spektrogramm $|X(\omega_k, \tau)|$ für ein Sprachsignal dargestellt. Werden reellwertige Signale untersucht, kann die Symmetrie der Fouriermatrix ausgenutzt werden, so dass lediglich

$$K = \frac{L_{\text{FFT}}}{2} + 1 \quad (4.11)$$

individuelle Frequenzbänder betrachtet werden müssen. In dem Beispiel aus Abbildung 4.3 (c) wurden $L_{\text{FFT}} = 512$ gewählt und die ersten 257 Frequenzbänder dargestellt.

Rücktransformation

Für die Rücktransformation werden die Inversen der beiden besprochenen Operationen durchgeführt. Die inverse diskrete Fouriertransformation (IDFT) ist durch

$$\tilde{x}_b(n, \tau) = \mathcal{F}^{-1} X(\omega_k, \tau) \quad (4.12)$$

gegeben. Danach kann das Signal stückweise zusammengesetzt werden

$$\tilde{x}(n) = \sum_{\tau=-\infty}^{\infty} \tilde{x}_b(n, \tau) * \delta(n - \tau L_V), \quad (4.13)$$

indem die Teilstücke entsprechend ihrer Verschiebung τL_V aufaddiert werden.

Rekonstruktionsbedingungen

Um eine korrekte Rücktransformation erreichen zu können, müssen $\gamma(n)$ und L_V der folgenden Bedingung genügen:

$$\sum_{\tau=-\infty}^{\infty} \gamma(n + \tau L_V) = \text{konst.} \quad (4.14)$$

Diese stellt sicher, dass die verschobenen Fenster sich zu einer Konstante addieren und somit alle Samples die gleiche Gewichtung bekommen. Die Konstante ist ein Verstärkungsfaktor, welcher durch eine einfache Multiplikation wieder herausgerechnet werden kann.

Die Bedingung (4.14) erfüllen alle Fenster, bei denen sich die erste und zweite Hälfte zu einer Konstanten mit

$$\gamma(n) + \gamma\left(n + \frac{L_\gamma}{2}\right) = \text{konst}, \quad 0 \leq n < \frac{L_\gamma}{2} \quad (4.15)$$

aufaddiert. Besonders gute spektrale Eigenschaften haben symmetrische Fenster mit

$$\gamma(n) = \gamma(L_\gamma - n), \quad 0 < n < L_\gamma. \quad (4.16)$$

Um die Korrektheit von (4.14) sicherzustellen, muss weiterhin der Vorschub

$$L_V = \frac{L_\gamma}{k}, \quad L_V, k \in \mathbb{N}, \quad k \geq 2 \quad (4.17)$$

ein ganzzahliger Bruchteil der Fensterlänge sein. Wird für (4.9) die FFT benutzt, so sind für L_γ und L_V Zweierpotenzen optimal.

Typische Fenster

Es existieren diverse Fenster, die den Rekonstruktionsbedingungen genügen. Sie haben unterschiedliche Eigenschaften im Frequenzbereich. Das *Rechteckfenster*

$$\gamma(n) = 1, \quad 0 \leq n < L_\gamma \quad (4.18)$$

hat den schmalsten Durchlassbereich aller Fenster, jedoch ist die Dämpfung der restlichen Frequenzen schwach. Die Blockbildung in der Gleichung (4.8) ist eine implizite Multiplikation mit einem Rechteckfenster.

Das *Hann-Fenster*

$$\gamma(n) = 0.5 - 0.5 \cos\left(\frac{2\pi n}{L_\gamma}\right), \quad 0 \leq n < L_\gamma, \quad (4.19)$$

4 Konvolutive Mischungen

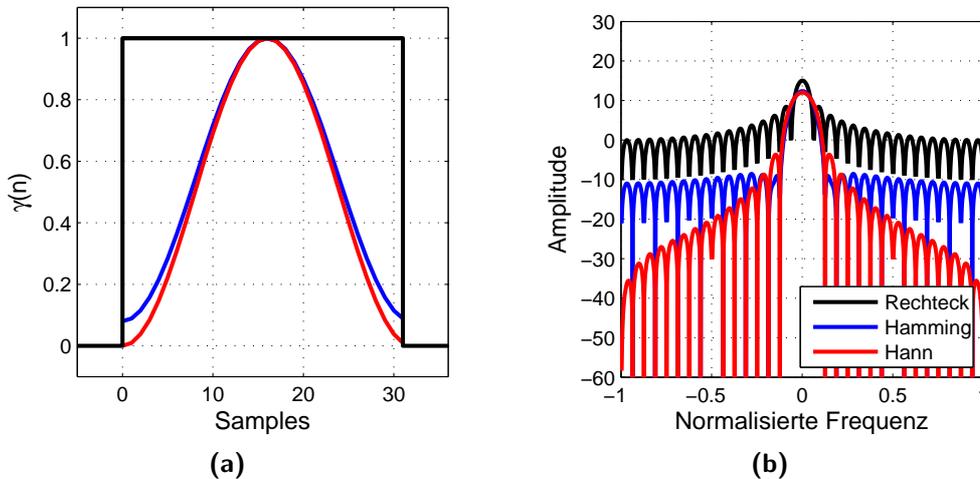


Abbildung 4.4: Übersicht über typische Fenster. (a) In der Zeitdarstellung mit $L_\gamma = 32$. (b) Der zugehörige Frequenzgang.

auch als Raised-Cosinus-Fenster bekannt, hat einen deutlich breiteren Durchlassbereich, sperrt aber andererseits die restlichen Frequenzen deutlich besser.

Das dritte häufig benutzte Fenster ist das *Hamming-Fenster*

$$\gamma(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{L_\gamma}\right), \quad 0 \leq n < L_\gamma, \quad (4.20)$$

welches ähnliche Eigenschaften wie das Hann-Fenster hat. Jedoch ist die Sperrung der Frequenzen wesentlich gleichmäßiger.

Ein Überblick über die Zeitdarstellung und das Frequenzverhalten ist in Abbildung 4.4 gegeben. Weitere Fenster und deren Eigenschaften finden sich in [39] und [65].

Filterung im Frequenzbereich

Die Transformation mit der STFT erlaubt eine einfache Verarbeitung der Signale im Frequenzbereich. Allerdings muss noch eine weitere Randbedingung erfüllt werden. Bei der Benutzung der DTFT führt eine Multiplikation im Frequenzbereich zu einer Faltung im Zeitbereich. Dies erlaubt eine direkte Verarbeitung der einzelnen Frequenzen.

Bei der DFT führt hingegen eine Multiplikation im Frequenzbereich zu einer zirkulären Faltung im Zeitbereich. Dieses Verhalten ist unerwünscht, kann je-

4.3 Konvolutives Misch- und Entmischungsmodell

doch durch geschickte Wahl der Parameter in eine lineare Faltung überführt werden.

Die zentrale Überlegung ist an dieser Stelle analog zum Verfahren der schnellen Faltung mittels der *Overlap-Add-Methode*. Hierbei wird die Faltung eines langen Signals $x(n)$ mit einem Filter $h(n)$, welche im Zeitbereich einen hohen Aufwand bedeutet, blockweise im Frequenzbereich mittels einer Multiplikation durchgeführt. Um das Problem der zirkulären Faltung zu umgehen, werden die einzelnen Blöcke so weit verlängert, bis die Faltung linear wird. Bei der Rücktransformation in den Zeitbereich werden die überlappenden Teile dann entsprechend aufaddiert. Wird nun ein Block $x_1(n)$ mit der Länge L_x mit dem Filter $h(n)$ mit der Länge L_h gefaltet, so folgt für das Ergebnis

$$y_1(n) = x_1(n) * h(n) \quad (4.21)$$

eine Länge von

$$L_y = L_x + L_h - 1. \quad (4.22)$$

Für eine schnelle Faltung im Frequenzbereich muss also die FFT-Länge mindestens L_y sein:

$$y_1(n) = \text{IDFT}[\text{DFT}(x_1(n), L_y) \cdot \text{DFT}(h(n), L_y)]. \quad (4.23)$$

$\text{DFT}(x(n), L_y)$ ist dabei DFT eines Signals $x(n)$, welches auf die Länge L_y mit Nullen aufgefüllt wurde. Der nächste Block $y_2(n)$ hat eine Überlappung mit den letzten L_h Werten aus $y_1(n)$ und muss entsprechend aufaddiert werden.

Gleiche Bedingungen gelten für die Verarbeitung mit Hilfe der STFT. Werden einzelne Frequenzbänder skaliert, so entspricht dies einer Faltung im Frequenzbereich. Um zirkuläre Artefakte zu vermeiden, muss sichergestellt sein, dass das zugehörige Filter die bei der Transformation vorgegebene Länge nicht überschreitet.

4.3 Konvolutives Misch- und Entmischungsmodell

Das konvolute Mischungsmodell wurde bereits in Gleichung (4.5) hergeleitet:

$$\mathbf{x}(n) = \mathbf{H}(n) * \mathbf{s}(n) = \sum_{l=0}^{L_H-1} \mathbf{H}(l) \mathbf{s}(n-l). \quad (4.24)$$

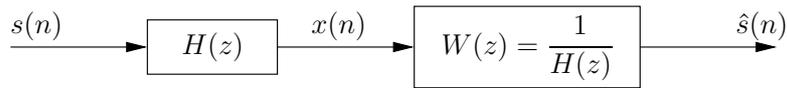


Abbildung 4.5: Berechnung des Inversen Systems mit Hilfe des Direct Inverse Equalizers.

Hierbei handelt es sich um ein *Multiple Input Multiple Output* (MIMO) System [28]. Um die gemischten Signale $\mathbf{x}(n)$ zu entmischen, muss also das inverse System $\mathbf{W}(n)$ gefunden werden. Doch zuerst müssen die Frage nach der Existenz von $\mathbf{W}(n)$ und gegebenenfalls dazu notwendige Bedingungen geklärt werden.

4.3.1 Existenz des inversen Systems

Im folgendem Abschnitt werden unterschiedliche Systeme nach der Existenz des inversen Systems untersucht. Angefangen wird mit einem SISO (Single Input Single Output) System. Dieses wird dann erweitert zu einem SIMO (Single Input Multiple Output) System. Abschließend wird das akustische Mischungssystem, welches ein MIMO (Multiple Input Multiple Output) System ist, beschrieben.

SISO Die Invertierung eines einzelnen Filters $h(n)$ wurde mehrfach untersucht [65, 28, 63]. Das inverse Filter $w(n)$ mit

$$h(n) * w(n) = \delta(n) \quad (4.25)$$

existiert, wenn $h(n)$ minimalphasig ist. Das heißt, dass alle Nullstellen von $H(z)$ innerhalb des Einheitskreises liegen. Hierbei ist $H(z)$ die z -Transformierte von $h(n)$ mit

$$H(z) = \sum_n h(n)z^{-n}. \quad (4.26)$$

Die Invertierung wird durchgeführt, indem alle Null- und Polstellen gegenseitig getauscht werden. Ein FIR-Filter wird damit zu einem IIR-Filter, dessen Ordnung unter Umständen sehr hoch und somit numerisch problematisch wird. Der *direct inverse (zero-forcing) equalizer*, wie in Abbildung 4.5 dargestellt, lautet in der z -Domäne

$$W(z) = \frac{1}{H(z)}. \quad (4.27)$$

Durch die Minimalphasigkeit von $h(n)$ wird die Stabilität von $w(n)$ sichergestellt. Raumimpulsantworten lassen sich im Allgemeinen nicht auf diese Art und Weise invertieren, da sie nicht minimalphasig sind [63].

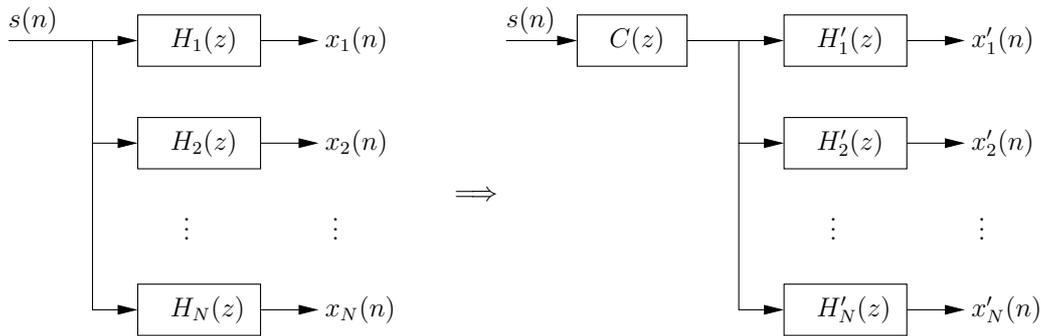


Abbildung 4.6: Reduktion eines SIMO Systems nach Gleichung (4.29)

SIMO Bei einem SIMO-System durchläuft ein Signal mehrere Filter $h_i(n)$ parallel. Der zentrale Ansatz ist die Unterschiedlichkeit aller Filter. Im Extremfall gleicher Filter würde sich das System zu einem SISO-System reduzieren. Bei einer Modellierung mit Hilfe von FIR-Filtern bedeutet Unterschiedlichkeit, dass die Systemfunktionen $H_i(z)$ unterschiedliche Nullstellen besitzen. Ein SIMO-System, dessen Filter keine gemeinsamen Nullstellen aufweisen, wird *nicht reduzierbar* genannt. Die Polynome $H_i(z)$ sind teilerfremd. Kann ein nicht teilerfremder Anteil

$$C(z) = \text{ggt}[H_1(z), H_2(z), \dots, H_N(z)] \neq 0 \quad (4.28)$$

mit Hilfe des größten gemeinsamen Teilers, $\text{ggt}(\cdot)$, abgespalten werden,

$$H_i(z) = C(z)H'_i(z), \quad (4.29)$$

so wird das System in einen SISO-Anteil $c(n)$ und ein nicht reduzierbares Restsystem $h'_i(n)$ aufgespalten. Diese Abspaltung ist in Abbildung 4.6 dargestellt. $c(n)$ kann nur wie ein SISO-System, mit all den verbundenen Nachteilen, invertiert werden, während für das Restsystem immer ein stabiles inverses System existiert. Dies basiert auf dem Bezout Theorem [58], welches besagt, dass für eine teilerfremde Menge an Polynomen $H_i(z)$ mit

$$\text{ggt}[H_1(z), H_2(z), \dots, H_N(z)] = 1 \quad (4.30)$$

immer eine Menge an Polynomen $W_i(z)$ existiert, so dass

$$\sum_{n=1}^N H_n(z)W_n(z) = 1 \quad (4.31)$$

gilt. Diese Beziehung ist auch als das *multichannel inverse theorem* (MINT) bekannt. Es besagt insbesondere, dass das inverse System mit FIR-Filtern erzeugt

4 Konvolute Mischungen

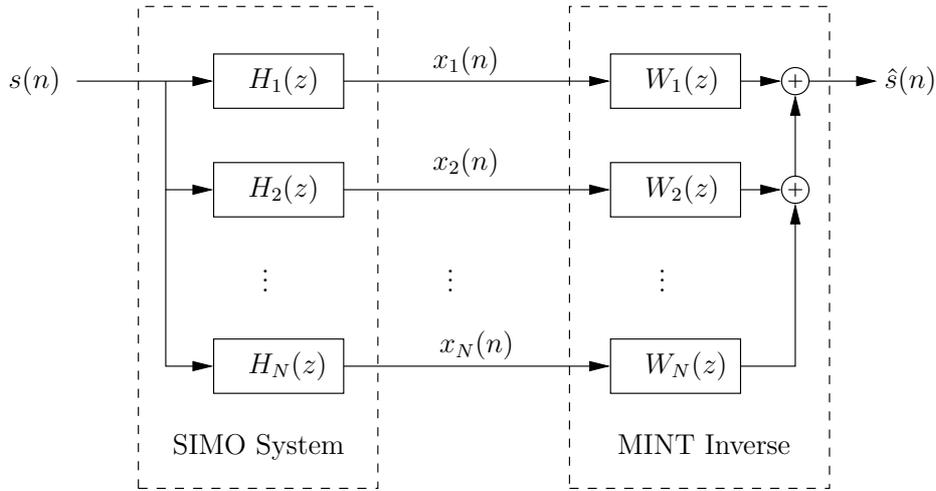


Abbildung 4.7: Die Berechnung des Inversen mit Hilfe des Mint Theorems

werden kann und somit auch immer stabil ist (Abbildung 4.7). Die Mindestlänge des inversen Systems ist mit

$$L_W = \left\lceil \frac{L-1}{N-1} \right\rceil \quad (4.32)$$

gegeben [28].

MIMO Hat das System mehrere Eingänge, wird das Problem deutlich komplizierter. Hier wird das System mit Hilfe von Impulsantworten $h_{ij}(n)$, welche die Übertragung von der j -ten Quelle zu dem i -ten Mikrofon angeben, beschrieben. Diese können dabei in einer Matrix-Folge $\mathbf{H}(n)$ zusammengefasst werden. Für dieses System kann ebenfalls das inverse System $\mathbf{W}(n)$ gefunden werden, sofern jedes Teilsystem, das zu einem Ausgang i gehört, nicht reduzierbar ist. Weiterhin muss gelten, dass die z-Transformierte $\mathbf{H}(z)$ invertierbar ist. In [71] wurde gezeigt, dass mit Filtern der Länge

$$L_W = (L-1)(N-1) \quad (4.33)$$

das inverse System realisierbar ist.

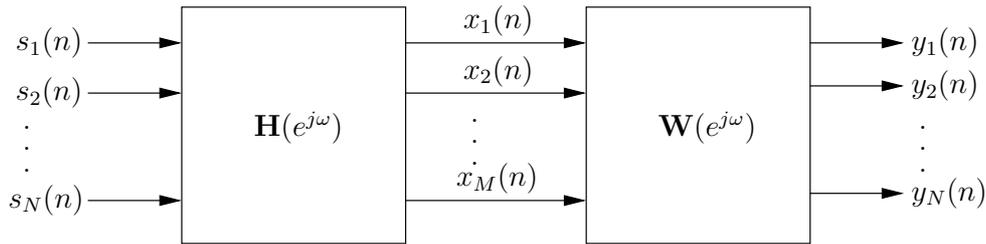


Abbildung 4.8: Das allgemeine System der konvolutiven blinden Quellentrennung. Sowohl das Mischungssystem $\mathbf{H}(e^{j\omega})$ wie auch das Entmischungssystem $\mathbf{W}(e^{j\omega})$ sind MIMO Systeme.

4.3.2 Das MIMO-Entmischungssystem

Das Mischungsmodell aus Gleichung (4.24) ist ein MIMO-System. Somit kann dafür, wie im letzten Abschnitt gezeigt, ebenfalls ein MIMO-Entmischungssystem

$$\mathbf{y}(n) = \mathbf{W}(n) * \mathbf{x}(n) = \sum_{l=0}^{L_W-1} \mathbf{W}(l) \mathbf{x}(n-l) \quad (4.34)$$

der Länge L_W gefunden werden. In Abbildung 4.8 ist dieses allgemeine Modell und in Abbildung 4.9 der Spezialfall eines 2×2 Misch- und Entmischungssystems abgebildet. In einem blinden Szenario sind dabei jedoch weder das Mischungssystem $\mathbf{H}(n)$ noch die Quellen $\mathbf{s}(n)$ bekannt. Somit können die bekannten Methoden für die Bestimmung des inversen Systems [28] nicht benutzt werden. Wenn aber von einer Unabhängigkeit der Quellen ausgegangen werden kann, so ist es möglich, mit Hilfe von statistischen Methoden ein Trennungssystem zu bestimmen. Die Formulierungen aus dem letzten Kapitel können so erweitert werden, dass auch konvolutive Mischungen behandelt werden können [25, 1]. In der Folge kann eine Kostenfunktion definiert werden, um die Koeffizienten direkt im Zeitbereich zu finden. Dieser Ansatz ist jedoch sehr aufwändig und bedarf einer sehr hohen Rechenleistung. Weiterhin bleibt das Gradientenverfahren oftmals in lokalen Minima gefangen, also kann in realen Szenarios keine gute Trennung erreicht werden.

In der Literatur wird aus den oben genannten Gründen häufig [32, 61, 73, 69, 33] ein anderer Ansatz im Frequenzbereich benutzt. Hierbei wird ausgenutzt, dass die Faltung im Zeitbereich zu einer Multiplikation im Frequenzbereich wird. Mit der STFT lautet das Mischungsmodell aus Gleichung (4.24)

$$\mathbf{X}(\omega_k, \tau) = \mathbf{H}(\omega_k) \mathbf{S}(\omega_k, \tau) \quad (4.35)$$

4 Konvolutive Mischungen

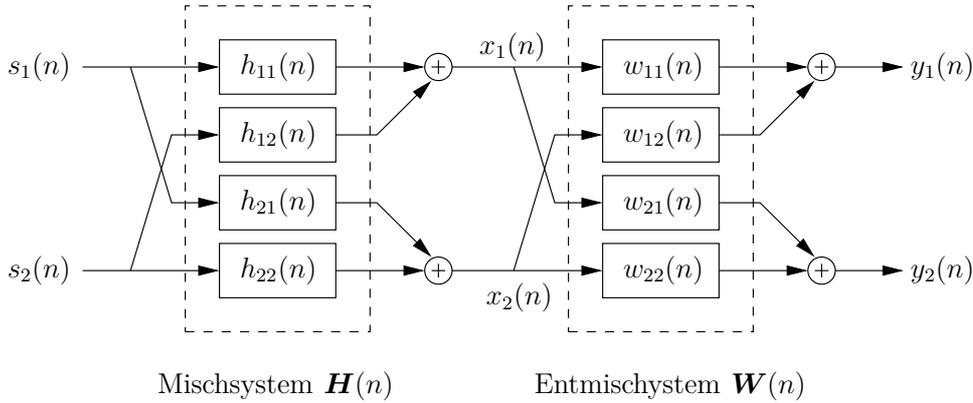


Abbildung 4.9: Der Spezialfall des (2×2) -Systems

und das Entmischungsmodell aus Gleichung (4.34)

$$\mathbf{Y}(\omega_k, \tau) = \mathbf{W}(\omega_k)\mathbf{X}(\omega_k, \tau). \quad (4.36)$$

Das Gesamtsystem $\mathbf{G}(\omega_k)$ lautet

$$\mathbf{G}(\omega_k) = \mathbf{W}(\omega_k)\mathbf{H}(\omega_k), \quad (4.37)$$

und damit gilt

$$\mathbf{Y}(\omega_k, \tau) = \mathbf{G}(\omega_k)\mathbf{S}(\omega_k, \tau). \quad (4.38)$$

Hierbei sind $\mathbf{H}(\omega_k)$ und $\mathbf{W}(\omega_k)$ die DFTs von $\mathbf{H}(n)$ und $\mathbf{W}(n)$. $\mathbf{G}(\omega_k)$ ist die DFT des gesamten Systems

$$\mathbf{G}(n) = \begin{pmatrix} g_{11}(n) & g_{12}(n) & \dots & g_{1N}(n) \\ g_{21}(n) & g_{22}(n) & \dots & g_{2N}(n) \\ \dots & \dots & \dots & \dots \\ g_{N1}(n) & g_{N2}(n) & \dots & g_{N,N}(n) \end{pmatrix}, \quad (4.39)$$

bestehend aus den Impulsantworten $g_{ij}(n)$, welche die Übertragung von der j -ten Quelle zum i -ten Ausgang des Entmischungssystems beschreiben. Diese Formulierung hat den entscheidenden Vorteil, dass in jedem Frequenzband ω_k eine instantane Mischung vorliegt. Somit kann auch mit Hilfe der Methoden aus dem letzten Kapitel eine instantane Entmischungsmatrix $\mathbf{W}(\omega_k)$ für jedes Frequenzband gefunden werden. Die Suche kann dabei unabhängig von allen anderen Entmischungsmatrizen durchgeführt werden, was den Aufwand erheblich reduziert.

4.3 Konvolutives Misch- und Entmischungsmodell

Die beschriebene Vorgehensweise hat auch Nachteile. Da jedes Frequenzband einzeln skaliert wird, ist nicht sichergestellt, dass die Bedingungen der Rücktransformation der STFT erfüllt sind. So kann es passieren, dass die zu den Frequenzkoeffizienten $W_{ij}(\omega_k)$ gehörenden Filter $w_{ij}(n)$ zu lang sind, so dass zirkuläre Artefakte in den rekonstruierten Signalen auftreten. Eine neue Methode zur Verringerung dieser Artefakte wird im nächsten Kapitel vorgestellt.

Ein anderes, viel schwerwiegenderes Problem ist in der ICA-Methode selbst begründet. Bei der Annahme der statistischen Unabhängigkeit kann weder auf die Reihenfolge, noch auf die Skalierung der Signale geschlossen werden. Damit kann jedes Frequenzband beliebig permutiert und skaliert sein. Selbst wenn die Entmischung perfekt funktioniert, ist das Ergebnis im Frequenzbereich eine beliebig permutierte und skalierte Version der Quellsignale. Mit dem Gesamtsystem

$$\mathbf{G}(\omega_k) = \mathbf{\Delta}(\omega_k)\mathbf{\Pi}(\omega_k) \quad (4.40)$$

gilt für die getrennten Signale

$$\mathbf{Y}(\omega_k, \tau) = \mathbf{W}(\omega_k)\mathbf{X}(\omega_k, \tau) = \mathbf{\Delta}(\omega_k)\mathbf{\Pi}(\omega_k)\mathbf{S}(\omega_k, \tau). \quad (4.41)$$

Hierbei ist $\mathbf{\Pi}(\omega_k)$ eine Permutationsmatrix und $\mathbf{\Delta}(\omega_k)$ eine Diagonalmatrix. Diese Unsicherheiten müssen korrigiert werden, bevor $\mathbf{Y}(\omega_k, \tau)$ wieder in den Zeitbereich transformiert wird. Eine fehlende Korrektur der Skalierung führt zu einer Filterung im Zeitbereich und somit zu einer unerwünschten Färbung der Signale.

Eine Permutation der einzelnen Frequenzbänder hat sogar noch schlimmere Auswirkungen auf das Zeitsignal. Selbst wenn alle Frequenzbänder gut getrennt sind, führt eine Permutation dazu, dass Teile der anderen Signale in dem jeweiligen Ausgang vorhanden sind. Die daraus rekonstruierten Zeitsignale sind immer noch gemischt.

Diese *inneren Skalierungs- und Permutationsprobleme* werden in den folgenden zwei Kapiteln noch genauer besprochen und es werden unterschiedliche Lösungsansätze vorgestellt. Die *äußeren Skalierungs- und Permutationsprobleme*, die zum Skalieren und Vertauschen der gesamten Signale führen, können mit den vorgestellten Methoden jedoch nicht gelöst werden und werden im Rahmen dieser Arbeit nicht weiter untersucht.

In Abbildung 4.10 ist das Entmischungssystem inklusive der Lösung des Permutations- und Skalierungsproblems für ein (2×2) -System dargestellt. Zuerst werden die zwei gemischten Signale $x_1(n)$ und $x_2(n)$ mit Hilfe der STFT in den Zeit-Frequenzbereich überführt. Danach werden die Frequenzbänder der STFT

4 Konvolutive Mischungen

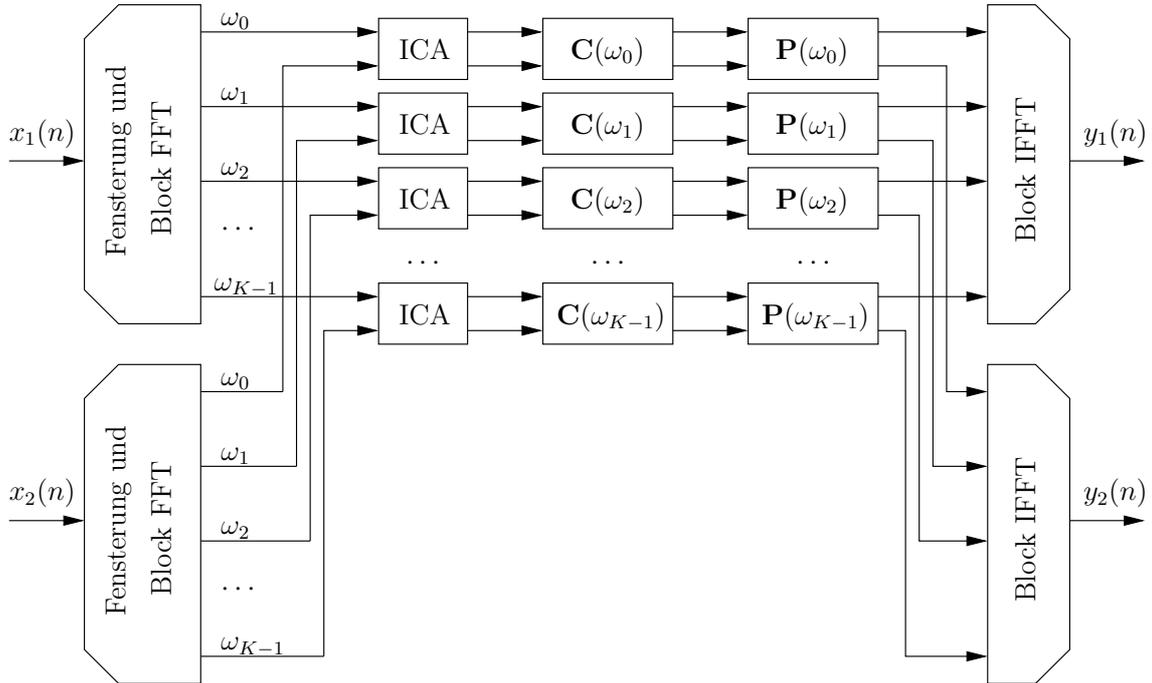


Abbildung 4.10: Das Entmischungssystem im Frequenzbereich mit der Korrektur der Permutation und Skalierung.

$\mathbf{X}(\omega_k, \tau)$ unabhängig voneinander mit Hilfe der instantanen ICA getrennt. Dem folgt die Korrektur der Skalierung mit Hilfe der Skalierungsmatrizen $\mathbf{C}(\omega_k)$ und die Lösung des Permutationsproblems mit $\mathbf{P}(\omega_k)$. Die so gewonnenen Signale $\mathbf{Y}(\omega_k, \tau)$ werden schließlich mit Hilfe der inversen STFT wieder in den Zeitbereich überführt.

Die in Abbildung 4.10 gewählte Reihenfolge der Lösung des Permutations- und Skalierungsproblems kann je nach gewählter Methode vertauscht werden, denn die dazu benutzten Verfahren basieren teilweise aufeinander. So gibt es zum Beispiel Methoden für die Lösung des Skalierungsproblems, die unabhängig von den benachbarten Frequenzbändern funktionieren [32, 46, 49]. Demgegenüber gibt es mit [51, 53] aber auch Methoden, die erst nach der Lösung des Permutationsproblems angewandt werden können.

Beispiel

Die einzelnen Schritte des Entmischungsalgorithmus sind in Abbildung 4.11 dargestellt. In Abbildung 4.11 (a) sind zwei Aufnahmen gemischter Signale in

4.3 Konvolutives Misch- und Entmischungsmodell

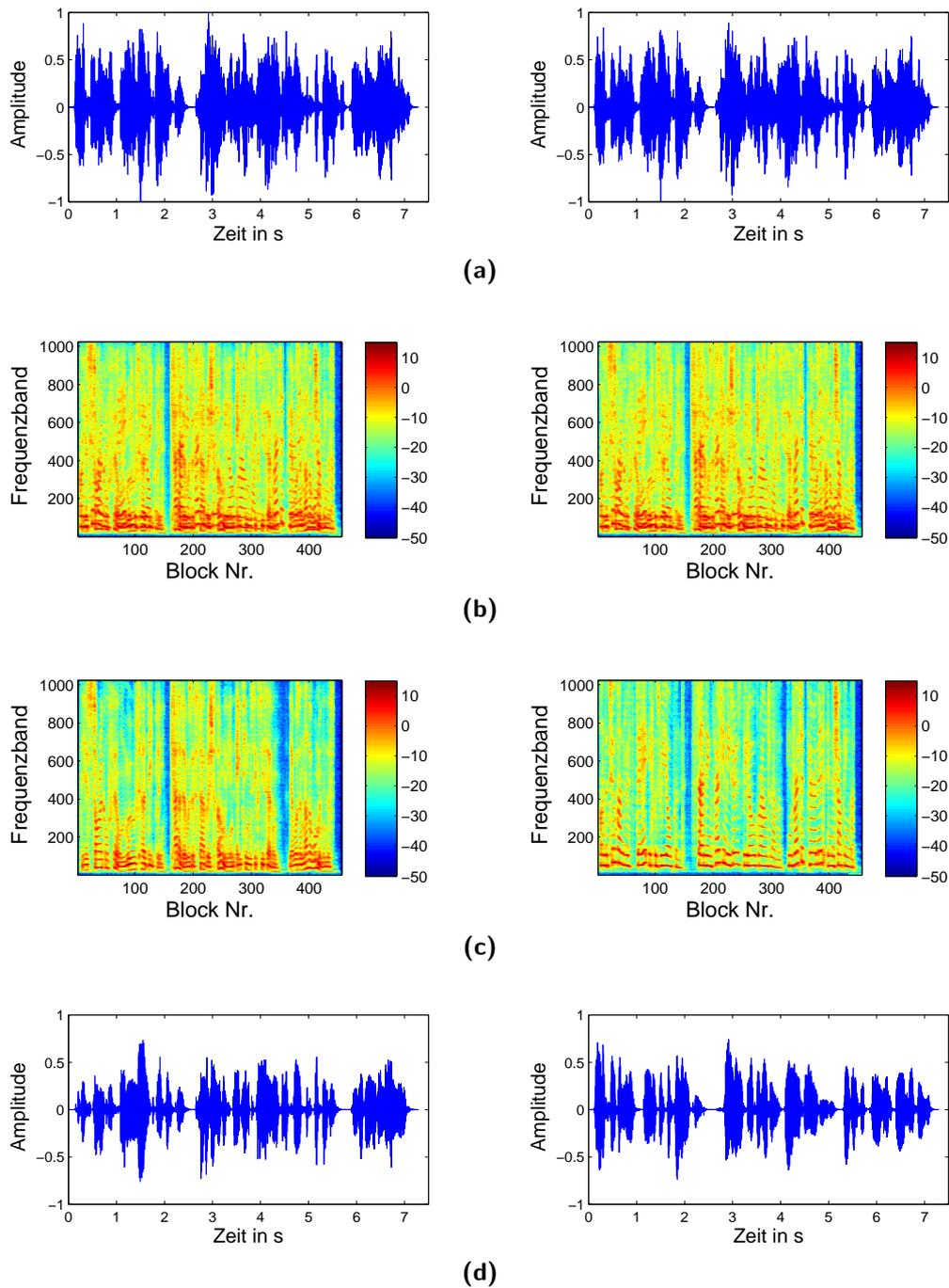


Abbildung 4.11: Die einzelnen Schritte der konvolutiven Trennung im Frequenzbereich. (a) Zwei Aufnahmen gemischter Signale in der Zeitdarstellung. (b) Die Aufnahmen in der Zeit-Frequenz Darstellung. (c) Die Zeit-Frequenz Darstellung nach der instantanen ICA in jedem Frequenzband. (d) Die getrennten Signale im Zeitbereich.

4 Konvolutive Mischungen

Zeitdarstellung gezeigt. In Abbildung 4.11 (b) sind die Signale nach der STFT gezeigt. Dann werden die Signale in jedem Frequenzband getrennt. Die so gewonnenen Spektrogramme sind in Abbildung 4.11 (c) dargestellt. Hier ist das Permutations- und Skalierungsproblem gelöst. Die Transformation in den Zeitbereich aus Abbildung 4.11 (d) liefert die korrekt getrennten Signale im Zeitbereich.

In Abbildung 4.12 wird beispielhaft gezeigt, was bei fehlender Korrektur mit den Signalen im Zeitbereich passiert. In Abbildung 4.12 (a) sind große Bereiche falsch permutiert. Die dazugehörigen Zeitsignale aus Abbildung 4.12 (b) sind nicht getrennt, obwohl die Trennung im Zeit-Frequenzbereich erfolgreich war. Eine falsche Skalierung aus Abbildung 4.12 (c) führt zu einer starken Filterung der Signale im Zeitbereich, wie in Abbildung 4.12 (d) dargestellt. Ohne eine Korrektur könnten die Signale bis zur Unkenntlichkeit verzerrt werden.

4.4 Bewertungsmethoden

In dem beschriebenen Verfahren zur konvolutiven Trennung im Frequenzbereich existieren mehrere Fehlerquellen, die die Trennleistung mindern. So kann unter anderem die instantane Trennung in den Frequenzbändern fehlschlagen, da entweder die Mischungsmatrix $\mathbf{H}(\omega_k)$ singulär oder nahezu singulär ist. Wird die Länge der Analysefenster L_γ oder der Entmischungsfiler L_W zu kurz gewählt, so ist auch die Annahme der instantanen Mischung und Entmischung in den Frequenzbändern nicht korrekt. Es kann auch passieren, dass die Permutationen nicht korrekt gelöst werden, was die Trennleistung erheblich mindern kann.

Die Bewertung der Trennleistung kann auf mehrere Arten durchgeführt werden. Die einfachste ist es, die Anzahl der falsch permutierten Frequenzbänder zu zählen. Dies ist keine verlässliche Methode, da je nach der Trennleistung sowie der Energie der betroffenen Bänder die Auswirkung einer falschen Permutation sehr unterschiedlich sein kann.

SIR

Um die Trennleistung des Systems zu beurteilen, kann das *signal-to-interference ratio* (Signal-zu-Störung Verhältnis, SIR) benutzt werden [74]. Dies kann auf der

4.4 Bewertungsmethoden

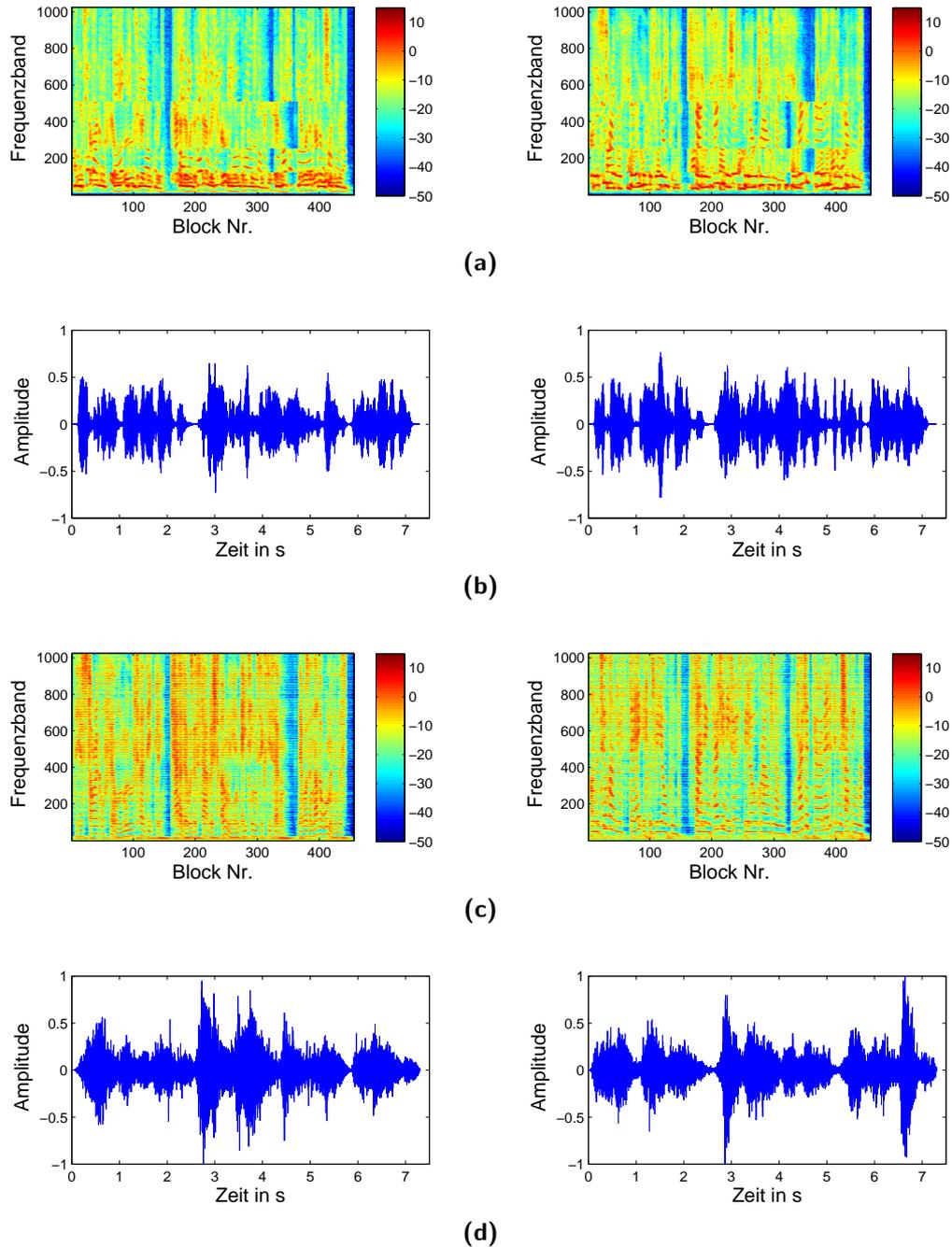


Abbildung 4.12: Beispiele für die fehlende Korrektur des Permutations- und Skalierungsproblems. (a) Das Permutationsproblem im Zeit-Frequenzbereich. (b) Die dazugehörigen Zeitsignale sind nicht getrennt. (c) Falsche Skalierung im Frequenzbereich. (d) Die Zeitsignale sind zwar getrennt, aber stark verzerrt.

4 Konvolutive Mischungen

Basis des Gesamtsystem $\mathbf{G}(n)$ für den i -ten Ausgang als

$$\text{SIR}_i = 10 \log_{10} \frac{E[(g_{ii}(n) * s_i(n))^2]}{E[(\sum_{j=1, j \neq i}^N g_{ij}(n) * s_j(n))^2]} \quad (4.42)$$

berechnet werden. Für die Berechnung des SIR ist dabei die Kenntnis über die Quellensignale oder das Mischsystem nicht notwendig. Es ist ausreichend, wenn die einzelnen Beiträge $x_{ij}(n)$ der Quellen $s_j(n)$ an den einzelnen Mikrofonen bekannt sind. Diese Möglichkeit vereinfacht den experimentellen Aufbau erheblich. Unter der Annahme der Linearität des akustischen Systems ist es ausreichend, die einzelnen Quellen nacheinander aufzunehmen und die gemischten Signale durch Addition der einzelnen Komponenten zu gewinnen. So kann die blinde Trennung auf diesen Signalen durchgeführt werden, während die Bewertung nicht blind mit Hilfe der Teilsignale erfolgen kann. Die in den weiteren Kapiteln beschriebenen Versuche werden auf so einem Datensatz [76] durchgeführt.

Das SIR kann auch für jedes Frequenzband einzeln bestimmt werden. Dies erlaubt insbesondere eine nicht blinde Bestimmung der Permutationen und kann als Referenz für die Qualität der Entpermutationsalgorithmen dienen.

SFM

Um die Skalierungseffekte, die zur Filterung der Signale führen, zu beurteilen, wird das *spectral flatness measure* (Spektrale Glattheit, SFM) benutzt [35, 47]. Die Glattheit eines Filters $w(n)$ wird auf den Frequenzkoeffizienten $W(\omega_k)$ mit

$$\text{SFM}_w = \frac{\sqrt{K} \prod_{k=0}^{K-1} |W(\omega_k)|^2}{\sum_{k=0}^{K-1} |W(\omega_k)|^2} \quad (4.43)$$

berechnet. Das SFM hat einen Wertebereich von $]0, 1]$. Ein Wert von eins bedeutet, dass das Filter ein Allpass ist und somit den Frequenzgang von $w(n)$ nicht verändert. Kleine Werte bedeuten größere Störung im Frequenzgang, und im Grenzfall von SFM gegen null wird nur eine Frequenz durchgelassen.

4.5 Zusammenfassung

In diesem Kapitel wurde das konvolutive Mischungs- und Entmischungssystem hergeleitet. Durch die Kurzzeit-Fouriertransformation wurden diese Modelle in

den Zeit-Frequenzbereich überführt, wo die Trennung instantan in allen Frequenzbändern unabhängig voneinander durchgeführt werden kann. Die dazu notwendigen Methoden wurden bereits im letzten Kapitel beschrieben.

Die Benutzung der ICA für die Trennung der einzelnen Frequenzbänder führt zu einem Skalierungs- und Permutationsproblem. Ohne eine Korrektur dieser Unsicherheiten sind die rekonstruierten Zeitsignale entweder stark gestört oder gar nicht getrennt. Methoden für die Lösung dieser Probleme werden in den nächsten zwei Kapiteln besprochen.

4 *Konvolute Mischungen*

Das Skalierungsproblem

Im vorausgegangenen Kapitel wurde gezeigt, wie die konvolutive blinde Quellentrennung durch die Transformation in den Zeit-Frequenzbereich in instantane Entmischungen überführt werden kann. Diese können dann unabhängig voneinander mit Hilfe der instantanen ICA getrennt werden. Eines der sich aus diesem Ansatz ergebenden Probleme ist die beliebige Skalierung jedes einzelnen Frequenzbandes. Diese Skalierung muss korrigiert werden, denn sonst wird lediglich eine unter Umständen sehr stark gefilterte Version der Quellen restauriert. Die unabhängige Verarbeitung in jedem Frequenzband kann weiterhin zur Folge haben, dass die Rekonstruktionsbedingungen der Kurzzeit-Fouriertransformation nicht eingehalten werden. Durch geschickte Wahl der Skalierungskoeffizienten können die Filter entsprechend geformt werden, um die zirkulären Artefakte zu minimieren.

In diesem Kapitel werden mehrere Methoden vorgestellt, mit denen die angesprochenen Probleme gelöst werden können. Die Lösungsansätze können in zwei Gruppen eingeteilt werden. Die erste Gruppe besteht aus Methoden, die für jede Frequenz unabhängig von allen anderen eine Skalierung bestimmen. Diese Methoden erlauben eine Lösung des Skalierungsproblems, ohne dass eine vorherige Lösung des Permutationsproblems benötigt wird. Das bekannteste Verfahren basiert auf inversen Nachfiltern [32] und versucht die Quellen so zu rekonstruieren, wie sie mit den Mikrofonen aufgenommen worden sind. Eine in diesem Kapitel vorgestellte und bereits in [49] veröffentlichte neue Methode versucht hingegen eine blinde Schätzung der Trennleistung in jedem Frequenzband. Dies kann dann genutzt werden, um die Trennleistung des Gesamtsystems zu verbessern.

Zu der zweiten Gruppe gehören Methoden, die erst eingesetzt werden können, wenn das Permutationsproblem gelöst ist. Da die Lösung des Permutationspro-

5 Das Skalierungsproblem

blems ihrerseits oftmals eine Lösung des Skalierungsproblem voraussetzt, ist die folgende Vorgehensweise möglich: Zuerst wird die Skalierung mit den inversen Nachfiltern durchgeführt, dann die korrekte Permutation bestimmt, und abschließend wird noch einmal skaliert. Bei gelöstem Skalierungsproblem können die Beziehungen zwischen den Koeffizienten benutzt werden, um die Form der Impulsantwort zu verändern. Hierzu werden zwei neue Methoden vorgestellt. Die erste Methode benutzt dazu die *Filter Verkürzung* mit dem Ziel, eine möglichst kurze Impulsantwort zu generieren [53]. Die zweite Methode benutzt die *Filterformung*, um exponentiell abfallende Impulsantworten zu formen [51], welche echten Raumimpulsantworten nachempfunden sind.

Problemdefinition

Die Lösung des Skalierungsproblems besteht aus reellen Skalierungskoeffizienten $c_i(\omega_k)$, die angeben, wie das k -te Frequenzband des i -ten Ausgangs verstärkt werden muss, damit das rekonstruierte Signal $y_i(n)$ eine möglichst gute Näherung an die unbekannte Quelle $s_i(n)$ ist.

Die Skalierungskoeffizienten $c_i(\omega_k)$ können entweder direkt mit den Frequenzbändern $Y_i(\omega_k, \tau)$ der getrennten Spektrogramme multipliziert werden oder durch das Zusammenfassen in einer Diagonalmatrix

$$\mathbf{C}(\omega_k) = \begin{bmatrix} c_{11}(\omega_k) & 0 & \dots & 0 \\ 0 & c_{22}(\omega_k) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & c_{N,N}(\omega_k) \end{bmatrix} \quad (5.1)$$

auf die Entmischungsmatrix

$$\mathbf{W}(\omega_k) = \begin{bmatrix} W_{11}(\omega_k) & W_{12}(\omega_k) & \dots & W_{1N}(\omega_k) \\ W_{21}(\omega_k) & W_{22}(\omega_k) & \dots & W_{2N}(\omega_k) \\ \dots & \dots & \dots & \dots \\ W_{N1}(\omega_k) & W_{N2}(\omega_k) & \dots & W_{N,N}(\omega_k) \end{bmatrix} \quad (5.2)$$

angewandt werden:

$$\mathbf{W}'(\omega_k) = \mathbf{C}(\omega_k)\mathbf{W}(\omega_k). \quad (5.3)$$

In dieser Schreibweise ist die i -te Zeile von $\mathbf{W}(\omega_k)$ ein Entmischungsvektor

$$\mathbf{w}_i(\omega_k) = [W_{i1}(\omega_k), W_{i2}(\omega_k), \dots, W_{iN}(\omega_k)]^T, \quad (5.4)$$

der die i -te Quelle extrahiert.

5.1 Einfache Lösungen

Gleiche Energie in allen Frequenzbändern

Es existieren zwei sehr einfache Methoden für die Lösung des Skalierungsproblems. Die erste Methode basiert auf der Annahme, dass die Energie des i -ten Signals

$$E_i(\omega_k) = \sum_{\tau=0}^{T-1} |\mathbf{Y}_i(\tau, \omega_k)|^2 \quad (5.5)$$

in allen Frequenzbändern ω_k konstant ist. Daraus ergeben sich direkt die Skalierungskoeffizienten

$$c_i(\omega_k) = \frac{1}{\sqrt{E_i(\omega_k)}}. \quad (5.6)$$

Dieser Ansatz eignet sich nicht für die Rekonstruktion von Sprachsignalen, da sich die Energie in den einzelnen Frequenzen sehr stark unterscheidet. Durch diesen Ansatz werden insbesondere die hohen Frequenzen verstärkt, die typischerweise eine um Größenordnungen kleinere Energie haben als die tiefen Frequenzen. In Abbildung 5.1 ist diese Verzerrung dargestellt. So zeigt die Zeit-Frequenz-Darstellung eines Sprachsignals aus Abbildung 5.1 (a) sehr große Energieunterschiede in den Frequenzbändern zu unterschiedlichen Zeitpunkten. Nach der Skalierung nach Gleichung 5.5 sind diese Unterschiede fast verschwunden.

Gleiche Länge aller Entmischungsvektoren

Bei der zweiten einfachen Methode werden keine Annahmen über die Signale gemacht. Für das Entmischungssystem werden Vektoren gleicher Länge benutzt. Hierzu werden die Skalierungskoeffizienten so gewählt, dass

$$\|\mathbf{w}'_i(\omega_k)\|_{\ell_2} = \left(\sum_{j=1}^N W_{ij}^2(\omega_k) \right)^{\frac{1}{2}} = 1 \quad (5.7)$$

gilt. Die unter dieser Annahme rekonstruierten Signale sind wesentlich näher an den originalen Quellen, wie der Vergleich der Abbildung 5.2 mit Abbildung 5.1 (a) zeigt. Hier wurden zwei Aufnahmen im Frequenzbereich getrennt und die Entmischungsmatrizen nach Gleichung (5.7) normalisiert. Die grobe Energieverteilung über der Frequenz ist zwar erhalten, aber der Vergleich mit Abbildung 5.1 (a) zeigt, dass die Energie in den meisten Frequenzbändern zu niedrig ist.

5 Das Skalierungsproblem

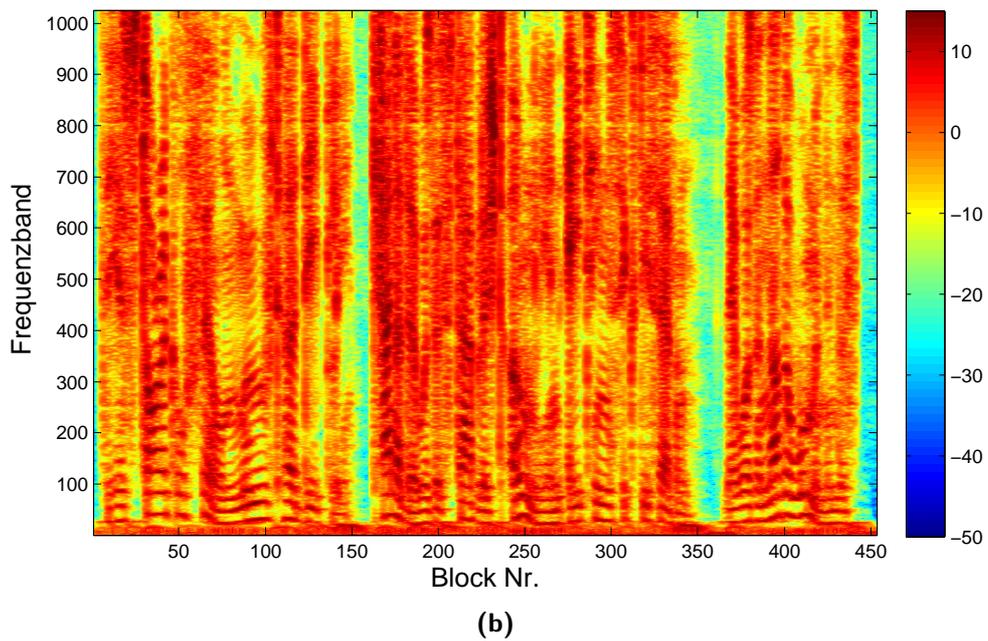
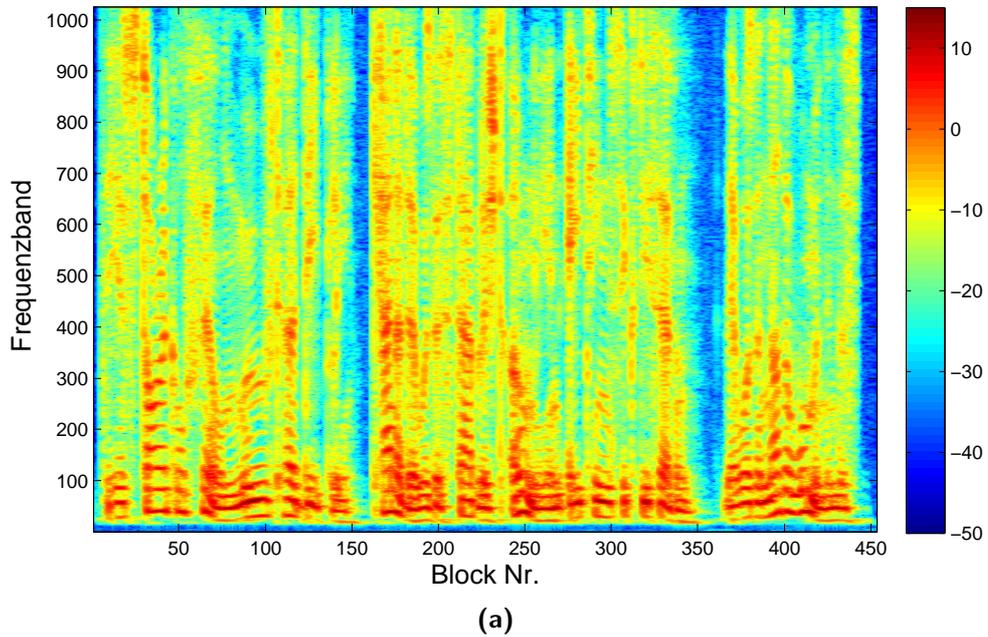


Abbildung 5.1: Die Lösung des Skalierungsproblems unter der Annahme gleicher Energie in allen Frequenzen. (a) Die Zeit-Frequenz-Darstellung eines Sprachsignals zeigt sehr große Energieunterschiede. (b) Entmischung unter der Annahme gleicher Energie in allen Frequenzbändern.

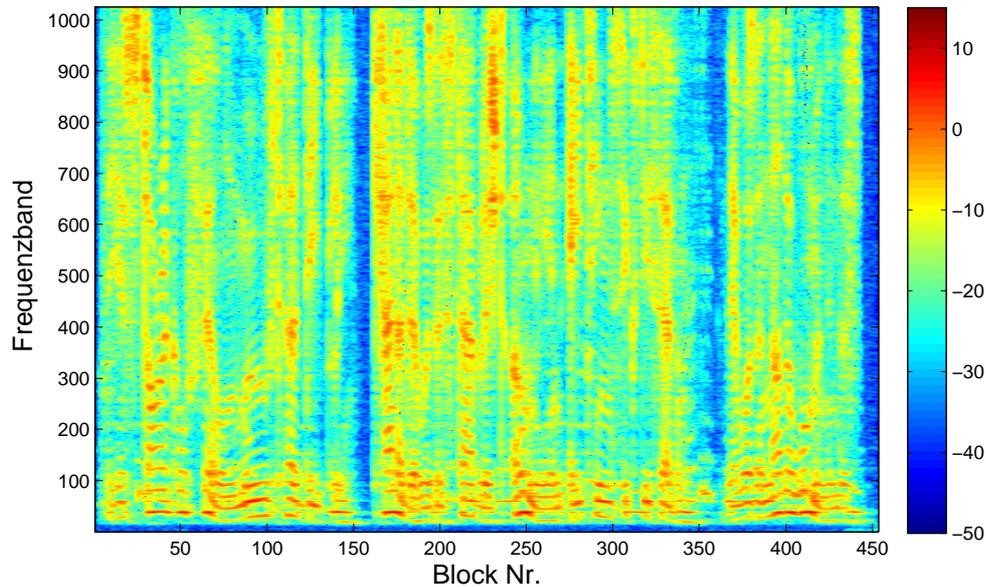


Abbildung 5.2: Die Lösung des Skalierungsproblems unter der Annahme gleicher Länge der Entmischungsvektoren. Die Energieverteilung über die Frequenzbänder ist grob erhalten, trotzdem haben, im Vergleich zu Abbildung 5.1 (a), viele Frequenzbänder eine zu kleine Energie.

Diese Verkleinerung der Energie ist nicht in allen Frequenzbändern gleich, so dass das rekonstruierte Signal entsprechend gefiltert ist.

5.2 Inverse Nachfilter

Um das Problem zu kleiner Energien in einzelnen Frequenzbändern zu lösen, wurde in [32] ein anderer Ansatz vorgestellt. Bei einem blinden Verfahren ist die Skalierung, die zu den originalen Signalen führt, grundsätzlich nicht ermittelbar. Aus diesen Gründen wurde in [32] vorgeschlagen, die Quellen so zu rekonstruieren, wie sie an den Mikrofonen aufgenommen worden sind. Dieser Ansatz akzeptiert die Filterung durch das Mischungssystem und fügt keine neuen Störungen hinzu. Dazu wurden die mit beliebiger Skalierung getrennten Frequenzbänder

$$\mathbf{Y}(\omega_k, \tau) = \mathbf{W}(\omega_k)\mathbf{X}(\omega_k, \tau) \quad (5.8)$$

5 Das Skalierungsproblem

mit Hilfe einer Selektionsmatrix \mathbf{S}^i einzeln ausgewählt und mit den Inversen $\mathbf{W}^{-1}(\omega_k)$ multipliziert:

$$\mathbf{Y}^i(\omega_k, \tau) = \mathbf{W}^{-1}(\omega_k) \mathbf{S}^i \mathbf{W}(\omega_k) \mathbf{X}(\omega_k, \tau). \quad (5.9)$$

Die Selektionsmatrix \mathbf{S}^i hat genau eine Eins an der i, i -ten Position und besteht ansonsten aus Nullen. So repräsentiert $\mathbf{S}^i \mathbf{W}(\omega_k) \mathbf{X}(\omega_k, \tau)$ genau einen Ausgang, so dass die Multiplikation mit dem Inversen der Entmischungsmatrix, also mit $\mathbf{W}^{-1}(\omega_k)$, die Bestimmung der einzelnen Teilkomponenten $x_{ij}(n)$ der aufgenommenen Signale $x_j(n)$ mit

$$x_j(n) = \sum_{i=1}^N x_{ij}(n) \quad (5.10)$$

erlaubt. Zu jeder Quelle werden N Teilkomponenten bestimmt, welche dann in $\mathbf{Y}^i(\omega_k, \tau)$ zusammengefasst sind. Werden für alle Quellen die Teilkomponenten bestimmt, so ergeben sich insgesamt N^2 Teilsignale.

Beispiel

Die Idee der inversen Nachfilter ist in Abbildung 5.3 dargestellt. Der Ausgangspunkt ist eine Mischung zweier Signale, die in der Abbildung 5.3 (a) zusammen mit den zugehörigen Entmischungsvektoren dargestellt ist. Die echten Komponenten sind unbekannt, jedoch liefert die ICA eine skalierte Version, wie in Abbildung 5.3 (b) abgebildet. In Abbildung 5.3 (c) wurden die getrennten Komponenten mit \mathbf{S}^i ausgewählt und jeweils einzeln farbig visualisiert. Schließlich ist in Abbildung 5.3 (d) die Anwendung der inversen Entmischungsmatrix auf die einzelnen Komponenten dargestellt. Aus dem Vergleich mit den gemischten Daten ist ersichtlich, dass die Skalierung der entmischten Daten genau der Skalierung der gemischten Daten entspricht. Das bedeutet, dass jegliche Skalierung nur aus dem unbekanntem Mischungssystem stammt, während das Entmischungssystem keine weitere Skalierung hinzufügt.

In den Abbildungen 5.4 und 5.5 ist der Unterschied zwischen der Entmischung mit inversen Nachfiltern und Vektoren konstanter Länge gegenübergestellt. Bei der Anwendung der inversen Nachfilter sind die Komponenten stets genau so wie in der Mischung skaliert. Dies gilt unabhängig von der Lage der Entmischungsvektoren. In dem Beispiel aus Abbildung 5.4 (a) müssen die Entmischungsvektoren entsprechend größer skaliert werden, während bei Abbildung 5.4 (b) die Länge nahezu unverändert bleibt. Dies ist gänzlich anders bei der Entmischung mit Vektoren konstanter Länge. Sind die Entmischungsvektoren senkrecht zueinander, wie in Abbildung 5.5 (b) dargestellt, so entspricht dies nahezu der Lösung

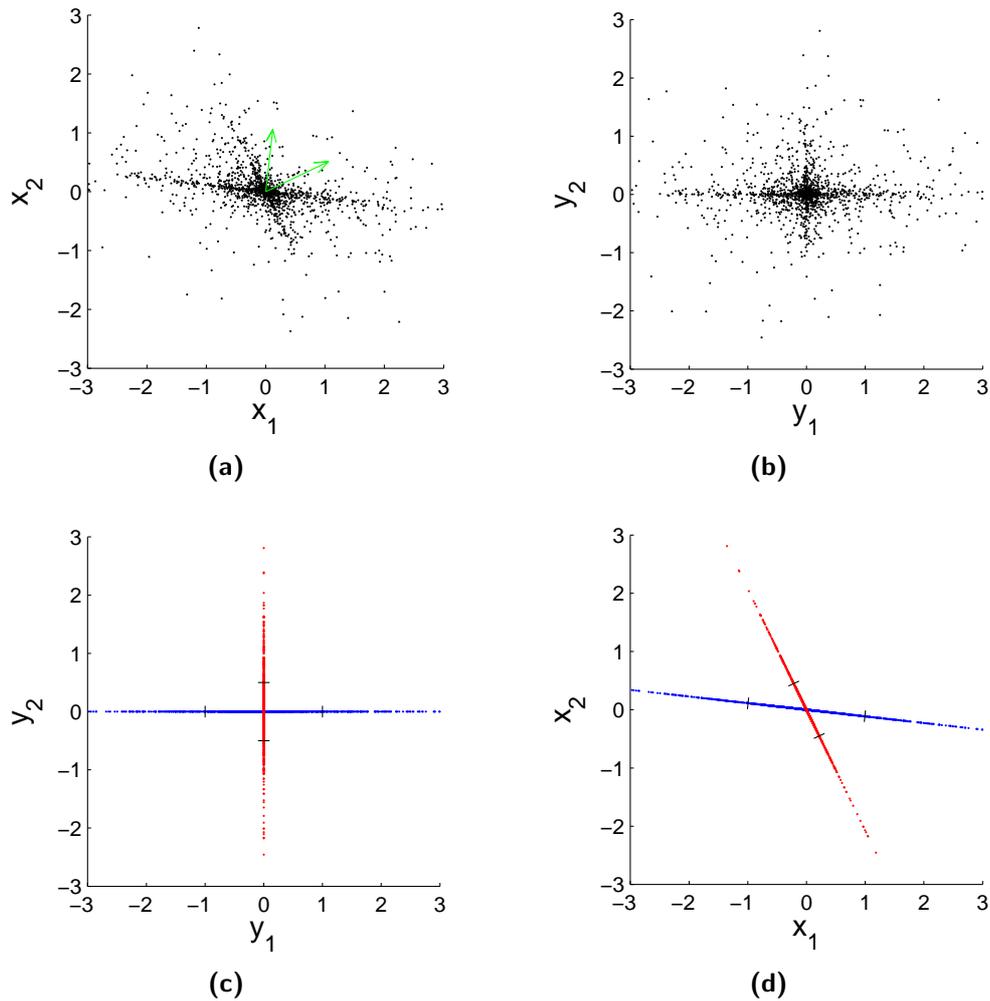


Abbildung 5.3: Die Visualisierung der Bestimmung der Komponenten $x_{ij}(n)$. (a) Mischung zweier Signale mit eingezeichneten Entmischungsvektoren $\mathbf{w}_1(\omega_k)$ und $\mathbf{w}_2(\omega_k)$ nach Gleichung (5.4). (b) Getrennte Daten. (c) Die beiden Ausgangssignale ausgewählt mit der Selektionsmatrix \mathbf{S}^i . (d) Die Anwendung der inversen Nachfilter auf die einzelnen Teilsignale.

mit den inversen Nachfiltern. Bei nahezu parallelen Entmischungsvektoren, wie in der Abbildung 5.5 (a) dargestellt, sind die getrennten Signale grundsätzlich zu klein skaliert.

5 Das Skalierungsproblem

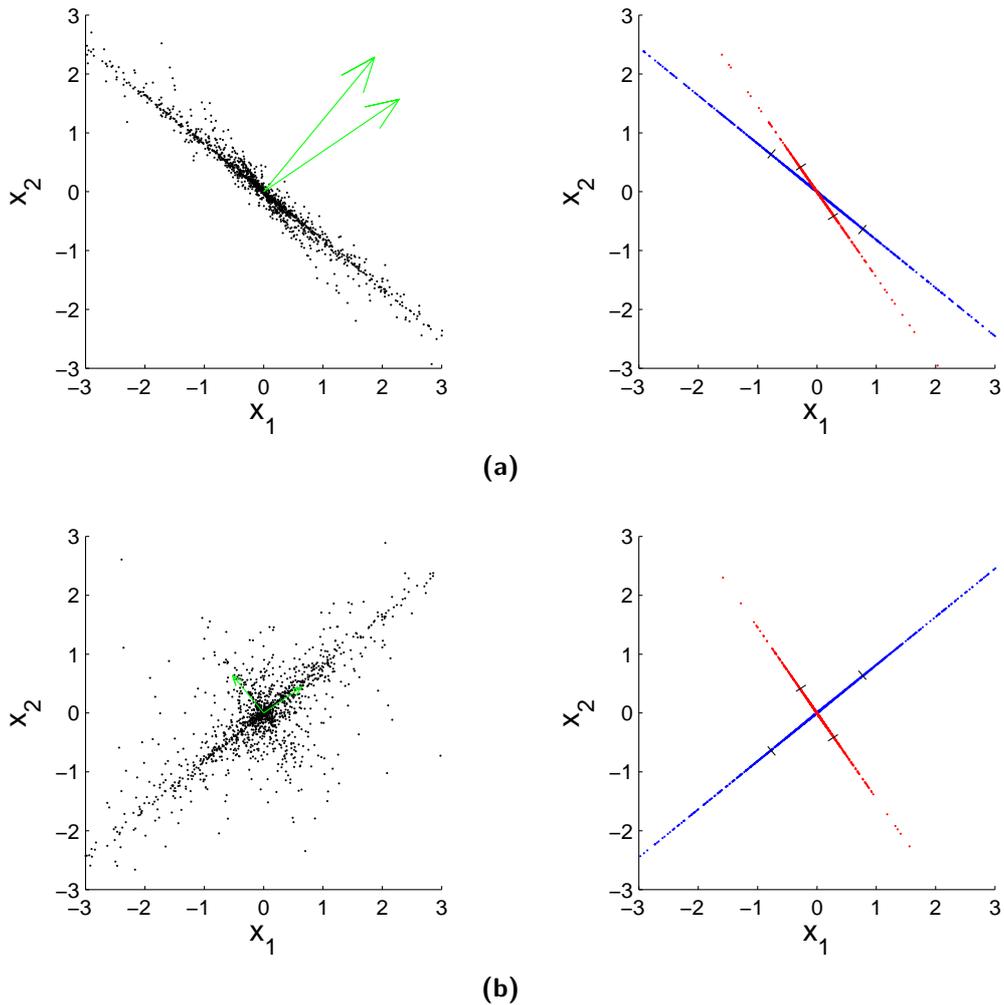


Abbildung 5.4: Die Skalierung mit inversen Nachfiltern. Die Entmischungsvektoren sind so skaliert, dass die getrennten Komponenten die gleiche Skalierung wie die Mischung haben. (a) Vektoren, die in eine ähnliche Richtung zeigen, werden verstärkt. (b) Bei nahezu senkrechten Vektoren ist eine sehr kleine Verstärkung nötig.

Minimal Distortion Principle

In [46] wurde ein ganz ähnliches Verfahren benutzt. Dort wurde das sogenannte *minimal distortion principle* (MDP) mit der Entmischungsmatrix

$$\mathbf{W}_{\text{mdp}}(\omega_k) = \text{ddiag}(\mathbf{W}^{-1}(\omega_k))\mathbf{W}(\omega_k) \quad (5.11)$$

vorge stellt. Hierbei ist $\text{ddiag}(\cdot)$ ein Operator, der alle Elemente einer Matrix, die nicht auf der Hauptdiagonalen liegen, auf Null setzt. Dieser Ansatz ist identisch

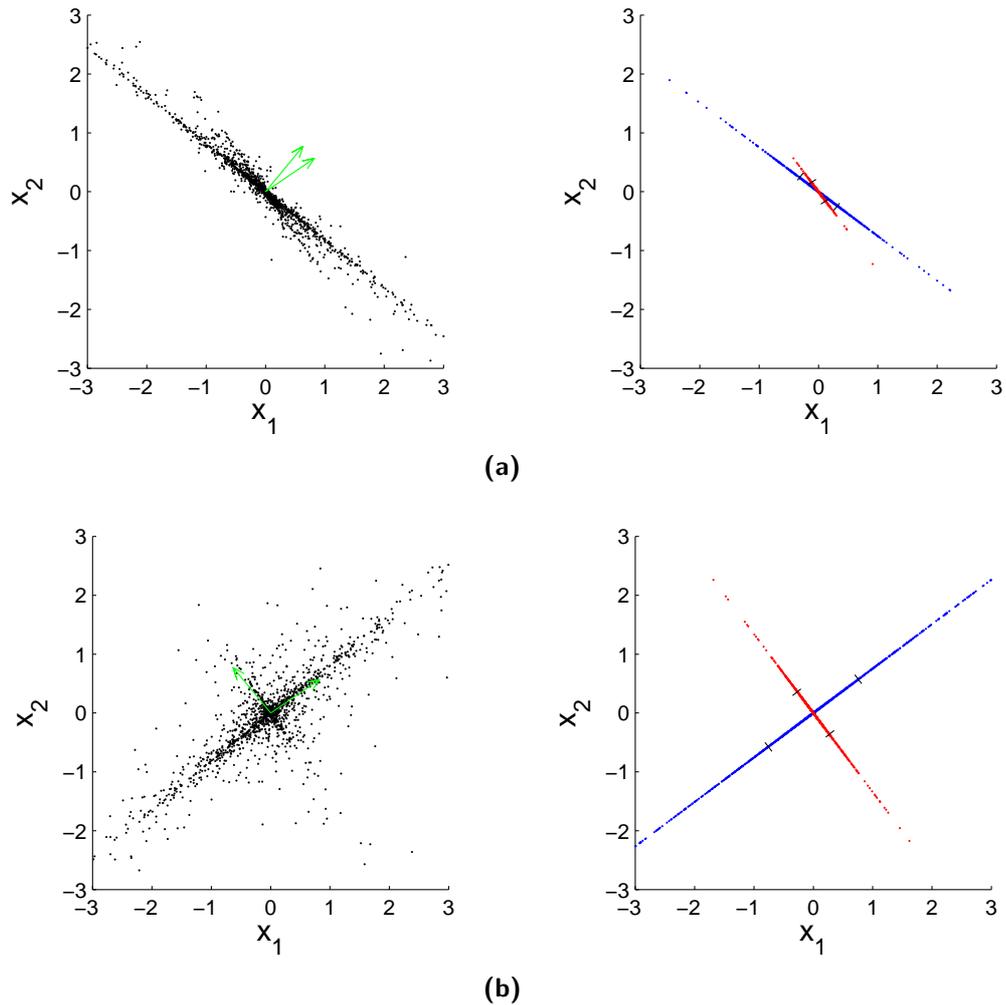


Abbildung 5.5: Die Skalierung unter der Annahme einer konstanten Länge der Entmischungsvektoren. (a) Bei Entmischungsvektoren, die in eine ähnliche Richtung zeigen, haben die rekonstruierten Signale eine zu kleine Energie. (b) Nur bei nahezu senkrechten Entmischungsvektoren ist die Energie der getrennten Signale erhalten.

mit der Nachfilterungsmethode aus Gleichung (5.9) unter der Vorgabe, dass die i -te Quelle am i -ten Sensor rekonstruiert wird. Das Ergebnis dieser Skalierung ist in Abbildung 5.6 dargestellt. Der Vergleich mit den Abbildungen 5.1 und 5.2 zeigt, dass die Energieverteilung mit Hilfe des MDP wesentlich besser rekonstruiert werden kann.

In [46] wurden auch weitere Eigenschaften von (5.11) gezeigt. So minimiert

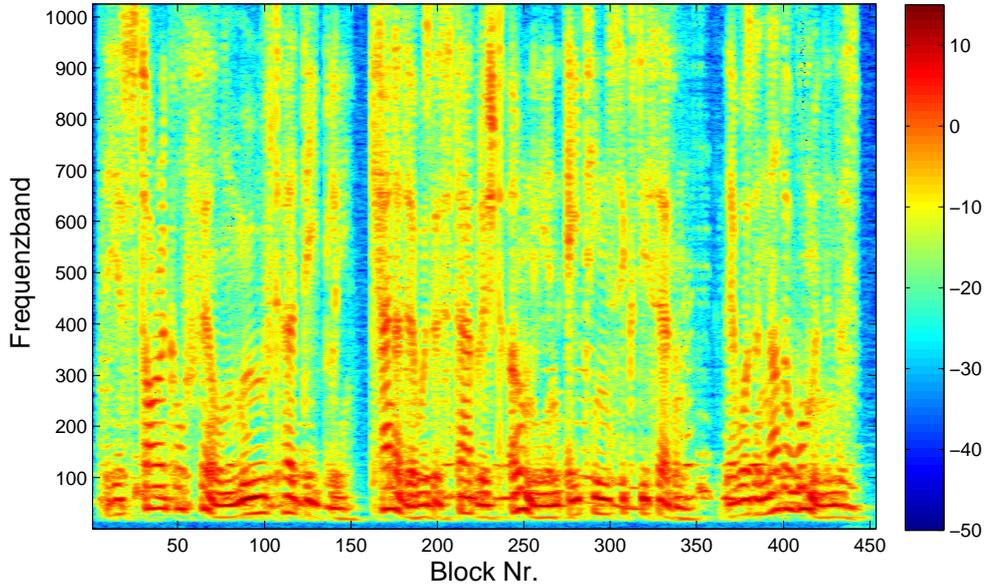


Abbildung 5.6: Die Lösung des Skalierungsproblems mit Hilfe des MDP.

$\mathbf{W}_{\text{mdp}}(\omega_k)$ den zu erwartenden Fehler

$$E\{|\mathbf{x}(n) - \mathbf{y}(n)|^2\} \quad (5.12)$$

und ist somit optimal in dem Sinne, dass es die getrennten Signale am wenigsten verzerrt.

Weiterhin ist mit dem *minimal distortion principle* auch

$$\|\mathbf{W}_{\text{mdp}}(\omega_k)\mathbf{H}(\omega_k) - \mathbf{H}(\omega_k)\|^2 \quad (5.13)$$

minimal, was bedeutet, dass jegliche Filterung lediglich vom Mischungssystem stammen kann. Im Weiteren wird $\mathbf{W}_{\text{mdp}}(\omega_k)$ als Basis und Referenz für weitere Algorithmen zur Korrektur der Skalierungsfaktoren benutzt.

5.3 Max-SIR Methode

In diesem Abschnitt wird eine neue Methode vorgestellt, die bereits in [49] veröffentlicht wurde. So wird im Folgenden untersucht, inwieweit die Trennleistung in jedem Frequenzband blind geschätzt werden kann. Ausgehend von dieser Schätzung werden dann die Skalierungskoeffizienten so bestimmt, dass die Trennleistung maximiert wird.

5.3.1 Blinde Bestimmung der Trennleistung

Die Bestimmung der Trennmatrix $\mathbf{W}(\omega_k)$ basiert auf einem endlichen Datensatz. Dies führt grundsätzlich zu einer fehlerbehafteten Schätzung. In Abhängigkeit von der Lage der Entmischungsvektoren $\mathbf{w}_i(\omega)$ zueinander führt dieser Fehler zu einem unterschiedlich starken Übersprechen. In dem Fall, dass die Mischungsvektoren, und damit auch die Entmischungsvektoren, nahezu parallel zueinander sind, ist dieser Fehler besonders problematisch, denn die Skalierung nach dem MDP verstärkt genau diese Frequenzen zusätzlich.

Die Position der Entmischungsvektoren zueinander kann mit dem Winkel α beschrieben werden. Für zwei komplexe Vektoren \mathbf{h}_1 und \mathbf{h}_2 gilt

$$\alpha = \arccos\left(\frac{|\mathbf{h}_1^H \mathbf{h}_2|}{\|\mathbf{h}_1\| \cdot \|\mathbf{h}_2\|}\right). \quad (5.14)$$

Auf diese Weise können nur Winkel kleiner als $\pi/2$ bestimmt werden. Dies ist aber kein Problem für die Schätzung, da größere Winkel auf

$$\alpha' = \pi - \alpha \quad (5.15)$$

abgebildet werden. Dies entspricht einer Skalierung eines der Vektoren mit dem Faktor $c = -1$ und hat keinen weiteren Einfluss auf die Trennung der Signale oder auf die Schätzung der Trennleistung.

Der ideale Fall

Der ideale Fall ist in Abbildung 5.7 dargestellt. Hier sind die Mischungsvektoren $\mathbf{h}_i(\omega_k)$ und Entmischungsvektoren $\mathbf{w}_i(\omega_k)$ biorthogonal zueinander. Damit lassen sich die Verstärkungsfaktoren

$$g_{ij}(\omega_k) = \mathbf{w}_i^H(\omega_k) \mathbf{h}_j(\omega_k) \quad (5.16)$$

von der Quelle j zum Ausgang i in einem 2×2 Fall als

$$\begin{aligned} g_{11}(\omega_k) &= \mathbf{w}_1^H(\omega_k) \mathbf{h}_1(\omega_k) \\ g_{12}(\omega_k) &= 0 \\ g_{21}(\omega_k) &= 0 \\ g_{22}(\omega_k) &= \mathbf{w}_2^H(\omega_k) \mathbf{h}_2(\omega_k) \end{aligned} \quad (5.17)$$

5 Das Skalierungsproblem

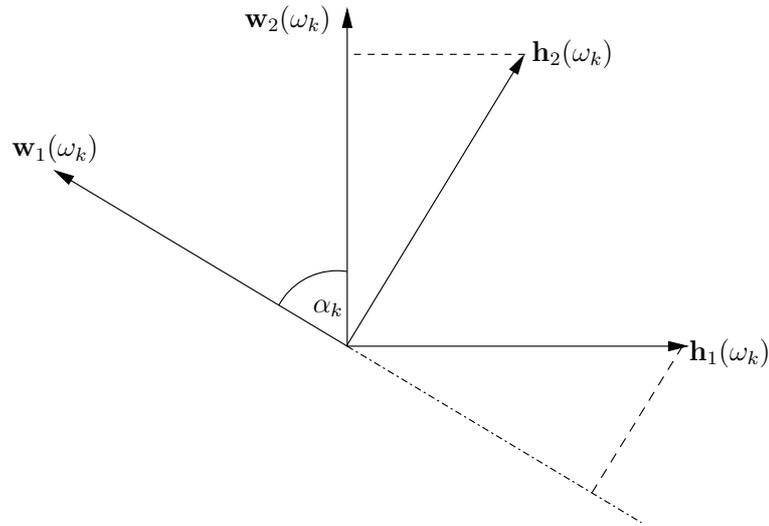


Abbildung 5.7: Projektion der Mischungs- und Entmischungsvektoren für die blinde Bestimmung der Trennleistung. Bei Einhaltung der Biorthogonalitätsbedingung sind die Signale perfekt getrennt.

beschreiben. Wegen der Orthogonalität

$$\mathbf{w}_1(\omega_k) \perp \mathbf{h}_2(\omega_k) \quad \text{und} \quad \mathbf{w}_2(\omega_k) \perp \mathbf{h}_1(\omega_k) \quad (5.18)$$

existiert kein Übersprechen, und die Signale sind perfekt getrennt. Mit Hilfe des MDP können die korrekten Skalierungen bestimmt werden. Die Position der Entmischungsvektoren zueinander, charakterisiert durch den Winkel α_k , spielt keine Rolle.

Der reale Fall

Der realistische Fall, bei dem die Biorthogonalitätsbedingung verletzt wird, ist in Abbildung 5.8 dargestellt. Aufgrund des endlichen Datensatzes werden die Entmischungsvektoren $\mathbf{w}_i(\omega_k)$ mit einem Fehler $\Delta \mathbf{w}_i(\omega_k)$ bestimmt:

$$\tilde{\mathbf{w}}_i(\omega_k) = \mathbf{w}_i(\omega_k) + \Delta \mathbf{w}_i(\omega_k) \quad (5.19)$$

Für die Verstärkungsfaktoren aus dem vorherigen (2×2) -Beispiel

$$g_{ij}(\omega_k) = \tilde{\mathbf{w}}_i^H(\omega_k) \mathbf{h}_j(\omega_k) \quad (5.20)$$

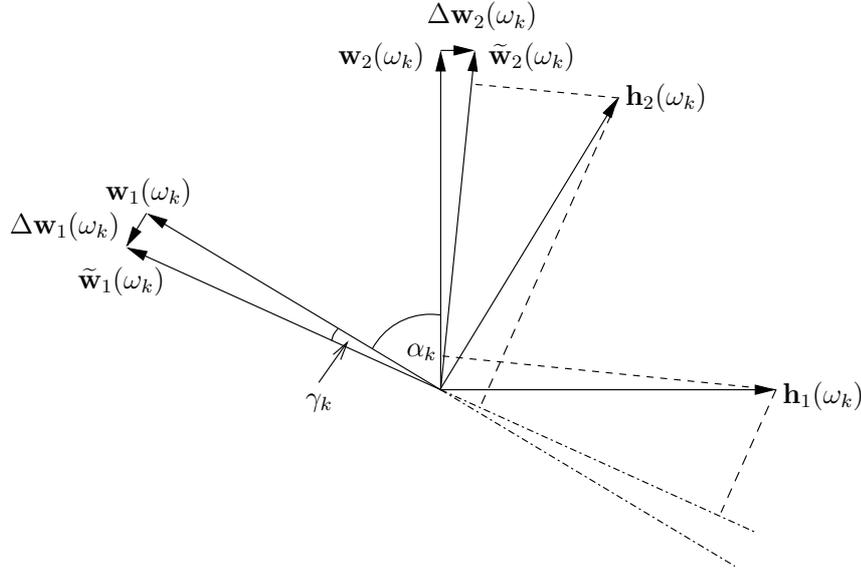


Abbildung 5.8: Die Projektion der Mischungsvektoren auf die Entmischungsvektoren bei der Verletzung der Biorthogonalitätsbedingung. Die Signale sind vollständig getrennt.

gilt dementsprechend

$$\begin{aligned}
 g_{11}(\omega_k) &= \mathbf{w}_1^H(\omega_k) \mathbf{h}_1(\omega_k) + \Delta \mathbf{w}_1^H(\omega_k) \mathbf{h}_1(\omega_k) \\
 g_{12}(\omega_k) &= \Delta \mathbf{w}_1^H(\omega_k) \mathbf{h}_2(\omega_k) \\
 g_{21}(\omega_k) &= \Delta \mathbf{w}_2^H(\omega_k) \mathbf{h}_1(\omega_k) \\
 g_{22}(\omega_k) &= \mathbf{w}_2^H(\omega_k) \mathbf{h}_2(\omega_k) + \Delta \mathbf{w}_2^H(\omega_k) \mathbf{h}_2(\omega_k).
 \end{aligned} \tag{5.21}$$

Da nun $g_{12}(\omega_k) \neq 0$ und $g_{21}(\omega_k) \neq 0$ gilt, existiert ein Übersprechen zwischen den Kanälen, und die Signale sind nicht mehr perfekt getrennt. Für die Trennleistung im ersten Ausgang gilt

$$\text{SIR}_1 = |S_1(\omega_k)|^2 \tag{5.22}$$

mit dem Übersprech-Faktor

$$S_1(\omega_k) = \frac{\mathbf{w}_1^H(\omega_k) \mathbf{h}_1(\omega_k) + \Delta \mathbf{w}_1^H(\omega_k) \mathbf{h}_1(\omega_k)}{\Delta \mathbf{w}_1^H(\omega_k) \mathbf{h}_2(\omega_k)}. \tag{5.23}$$

Analog kann für den zweiten Ausgang

$$\text{SIR}_2 = |S_2(\omega_k)|^2 \tag{5.24}$$

mit

$$S_2(\omega_k) = \frac{\mathbf{w}_2^H(\omega_k) \mathbf{h}_2(\omega_k) + \Delta \mathbf{w}_2^H(\omega_k) \mathbf{h}_2(\omega_k)}{\Delta \mathbf{w}_2^H(\omega_k) \mathbf{h}_1(\omega_k)} \tag{5.25}$$

5 Das Skalierungsproblem

definiert werden.

Im blinden Fall sind die Mischungsvektoren $\mathbf{h}_i(\omega_k)$ und die Fehler $\Delta\mathbf{w}_i(\omega_k)$ in den Gleichungen (5.23) und (5.25) nicht bekannt. Es ist jedoch möglich, eine Abschätzung für den schlimmsten Fall zu machen. Das Übersprechen ist maximal, wenn der Fehler $\Delta\mathbf{w}_1(\omega_k)$ orthogonal zum Entmischungsvektor $\mathbf{w}_1(\omega_k)$ mit

$$\mathbf{w}_1(\omega_k) \perp \Delta\mathbf{w}_1(\omega_k) \quad (5.26)$$

und kollinear zu $\mathbf{w}_2(\omega_k)$ mit

$$\Delta\mathbf{w}_1(\omega_k) \parallel \mathbf{w}_2(\omega_k) \quad (5.27)$$

ist. Weiterhin ist der Fehler $\Delta\mathbf{w}_1(\omega_k)$ sehr klein gegenüber $\mathbf{w}_1(\omega_k)$ und kann mit Hilfe des Winkels γ_k beschrieben werden:

$$\|\Delta\mathbf{w}_1(\omega_k)\| = \|\mathbf{w}_1(\omega_k)\| \sin(\gamma_k) \approx \|\mathbf{w}_1(\omega_k)\| \gamma_k. \quad (5.28)$$

Für den maximalen Fehler kann angenommen werden, dass

$$\Delta\mathbf{w}_1(\omega_k) \parallel \mathbf{h}_2(\omega_k) \quad (5.29)$$

ist und damit

$$|\Delta\mathbf{w}_1^H(\omega_k)\mathbf{h}_2(\omega_k)| = \|\Delta\mathbf{w}_1(\omega_k)\| \cdot \|\mathbf{h}_2(\omega_k)\| \quad (5.30)$$

gilt. Da die Mischungsvektoren nicht bekannt sind, ist die beste Annahme die Gleichheit der Längen

$$\|\mathbf{h}_1(\omega_k)\| = \|\mathbf{h}_2(\omega_k)\|. \quad (5.31)$$

Die letzten zu treffenden Annahmen sind, dass der Winkel α_k zwischen den Mischungsvektoren gleich dem Winkel zwischen den Entmischungsvektoren ist,

$$\alpha_k = \angle(\mathbf{h}_1(\omega_k), \mathbf{h}_2(\omega_k)) = \angle(\mathbf{w}_1(\omega_k), \mathbf{w}_2(\omega_k)), \quad (5.32)$$

und dass die Lage der Vektoren zueinander mit

$$\angle(\mathbf{w}_1(\omega_k), \mathbf{h}_1(\omega_k)) = \frac{\pi}{2} - \alpha_k \quad (5.33)$$

beschrieben werden kann. Für die Gültigkeit von (5.33) muss gegebenenfalls eine Skalierung nach Gleichung (5.15) durchgeführt werden.

Mit diesen Annahmen kann $|S_1(\omega_k)|$ blind wie folgt abgeschätzt werden:

$$\begin{aligned}
|S_1(\omega_k)| &\approx \frac{|\mathbf{w}_1^H(\omega_k)\mathbf{h}_1(\omega_k)| + |\Delta\mathbf{w}_1^H(\omega_k)\mathbf{h}_1(\omega_k)|}{|\Delta\mathbf{w}_1^H(\omega_k)\mathbf{h}_2(\omega_k)|} \\
&= \frac{\|\mathbf{w}_1(\omega_k)\| \cdot \|\mathbf{h}_1(\omega_k)\| \cos(\frac{\pi}{2} - \alpha_k)}{\|\Delta\mathbf{w}_1(\omega_k)\| \cdot \|\mathbf{h}_2(\omega_k)\|} \\
&\quad + \frac{\|\Delta\mathbf{w}_1(\omega_k)\| \cdot \|\mathbf{h}_1(\omega_k)\| \cos(\alpha_k)}{\|\Delta\mathbf{w}_1(\omega_k)\| \cdot \|\mathbf{h}_2(\omega_k)\|} \\
&\approx \frac{\|\mathbf{w}_1(\omega_k)\| \cos(\frac{\pi}{2} - \alpha_k)}{\|\Delta\mathbf{w}_1(\omega_k)\|} + \cos(\alpha_k) \\
&\approx \frac{\|\mathbf{w}_1(\omega_k)\| \cos(\frac{\pi}{2} - \alpha_k)}{\|\mathbf{w}_1(\omega_k)\| \gamma_k} + \cos(\alpha_k) \\
&= \frac{\cos(\frac{\pi}{2} - \alpha_k)}{\gamma_k} + \cos(\alpha_k).
\end{aligned} \tag{5.34}$$

Bei der Berechnung von $|S_2(\omega_k)|$ werden äquivalente Annahmen gemacht, so dass das Übersprechen in beiden Richtungen und damit auch die Trennleistung für beide Ausgänge gleich ist:

$$|S_1(\omega_k)|^2 = |S_2(\omega_k)|^2. \tag{5.35}$$

Der Winkel α_k kann mit Hilfe der Trennmatrix $\mathbf{W}(\omega_k)$ bestimmt werden. Der Fehler $\Delta\mathbf{W}(\omega_k)$ und der dazugehörige Winkel γ_k sind nicht bekannt, können aber als klein und konstant über der Frequenz angenommen werden. In der Praxis hat sich ein Wert von

$$\gamma_k = 0.001\pi \tag{5.36}$$

als realistisch herausgestellt [49]. Eine grobe Abschätzung von γ_k ist ausreichend, da durch die Normierung, die bei der Bestimmung der Skalierungskoeffizienten im nächsten Abschnitt durchgeführt wird, der Wert γ_k herausgekürzt wird.

Beispiel

Das vorgeschlagene Modell wurde auf realen Aufnahmen überprüft. In Abbildung 5.9 (a) ist die geschätzte Trennleistung nach dem vorgeschlagenen Modell berechnet. Demgegenüber ist in Abbildung 5.9 (b) die wahre Trennleistung dargestellt, welche nicht-blind aus den getrennten Daten bestimmt wurde.

5 Das Skalierungsproblem

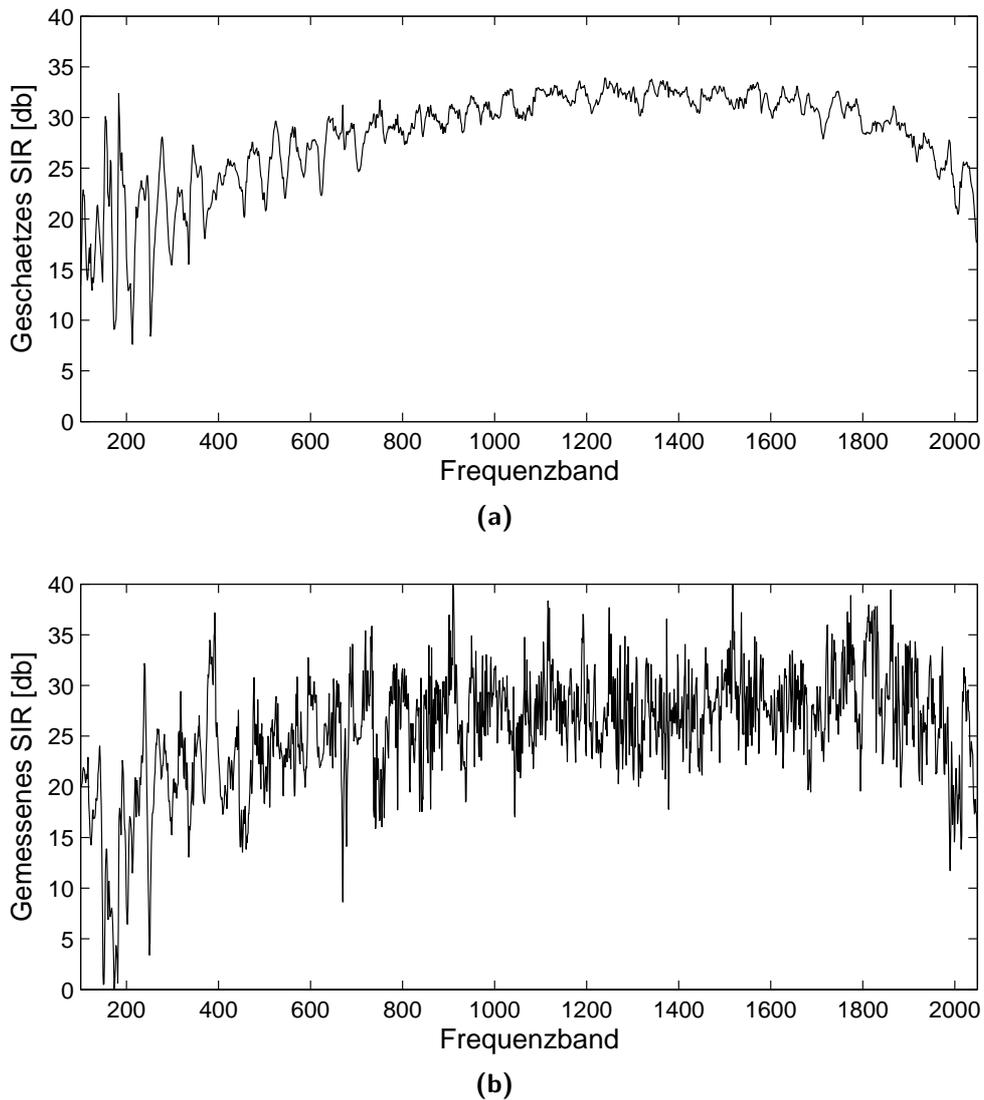


Abbildung 5.9: Beispiel für die Bestimmung der Trennleistung. (a) Die blind geschätzte Trennleistung nach Gleichung (5.22) und (b) die wahre Trennleistung.

Die beiden Graphen sind zwar nicht genau übereinstimmend, jedoch ist eine hohe Ähnlichkeit zu erkennen. Die niedrigen Frequenzen sind nicht so gut getrennt, was auch vom Modell vorhergesagt wird. Die mittleren und hohen Frequenzbereiche sind hingegen gut getrennt, was wiederum mit dem Modell übereinstimmt. Das Streubild der Daten in Abbildung 5.10 bestätigt diese hohe Ähnlichkeit. Die hohe Dichte an Punkten im oberen rechten Bereich bedeutet eine hohe Trennleistung für die meisten Frequenzbänder, welche auch vom Modell vorhergesagt

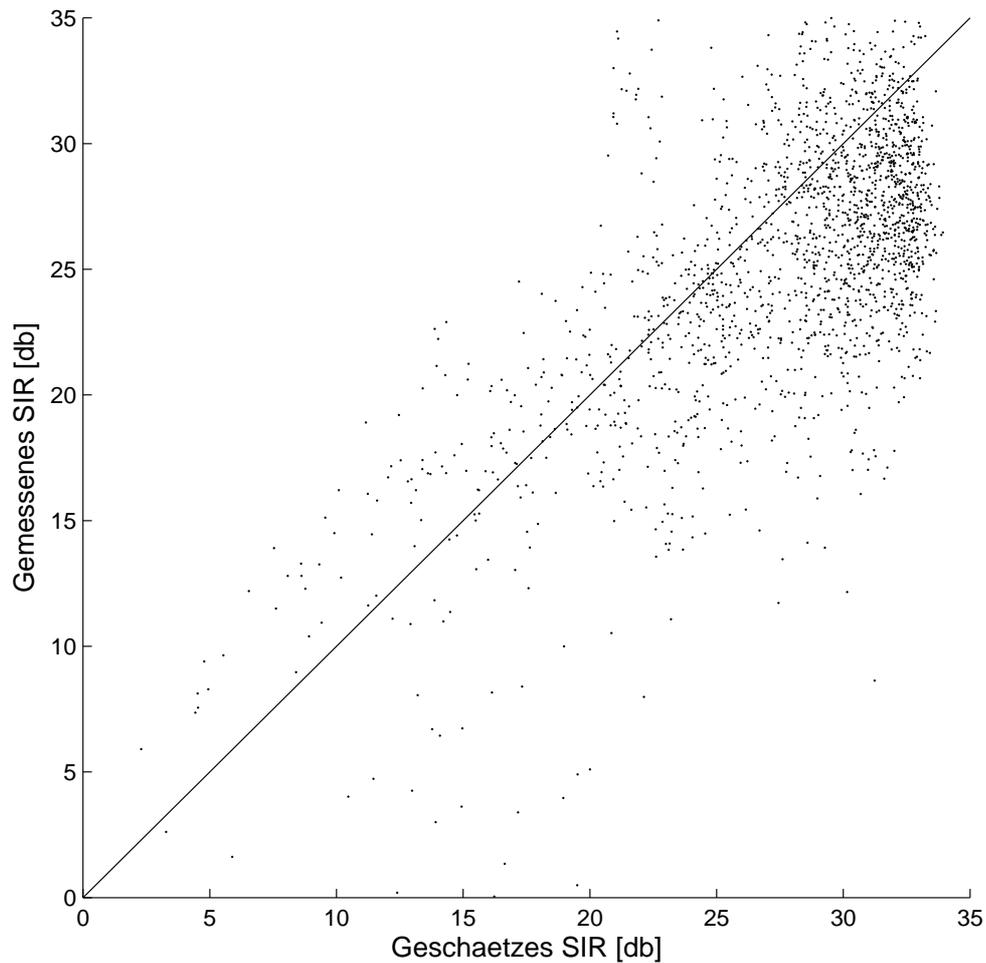


Abbildung 5.10: Der Scatterplot der geschätzten Trennleistung gegenüber der wahren Trennleistung. Es ist eine hohe Übereinstimmung zu erkennen.

wird. Auch die Schätzung der weniger gut getrennten Bänder ist gut gelungen, was an den Punkten entlang der Diagonalen sichtbar ist. Die hohe Übereinstimmung wird durch den Korrelationskoeffizienten von

$$r = 0.62 \quad (5.37)$$

bestätigt. Insgesamt lässt sich die Trennleistung mit dem vorgeschlagenen Modell sehr gut blind abschätzen.

5.3.2 Berechnung der Skalierungskoeffizienten

Ausgehend von der geschätzten Trennleistung können nun die Skalierungskoeffizienten so bestimmt werden, dass die gesamte Trennleistung der rekonstruierten Signale maximiert wird. Dies wird genau dann erreicht, wenn die Energie des Übersprechens in allen Frequenzbändern gleich groß ist. Für die Bestimmung der Skalierungskoeffizienten sind die absoluten Größen nebensächlich. Mit der Normierung der Übersprech-Faktoren

$$\tilde{S}_i(\omega_k) = \frac{|S_i(\omega_k)|}{\frac{1}{k} \sum_{k=0}^{K-1} |S_i(\omega_k)|} \quad (5.38)$$

und der Energie der getrennten Signale nach Gleichung (5.5),

$$\tilde{E}_i(\omega_k) = \frac{|E_i(\omega_k)|}{\frac{1}{k} \sum_{k=0}^{K-1} |E_i(\omega_k)|}, \quad (5.39)$$

können die Skalierungskoeffizienten als

$$c_1(\omega_k) = \tilde{S}_1(\omega_k) \cdot \tilde{E}_2(\omega_k) \quad (5.40)$$

und

$$c_2(\omega_k) = \tilde{S}_2(\omega_k) \cdot \tilde{E}_1(\omega_k) \quad (5.41)$$

berechnet werden. Für die Berechnung der Energie $\tilde{E}_i(\omega_k)$ der Frequenzbänder in Gleichung (5.39) kann dabei die Skalierung nach dem MDP vorgenommen werden. Die Koeffizienten $c_i(\omega_k)$ sind dann als weitere Skalierung zu verstehen und geben an, wie die Frequenzbänder gegenüber dem MDP verstärkt werden.

Die Skalierungskoeffizienten aus Gleichung (5.40) und (5.42) maximieren zwar die Trennleistung, führen aber auch gleichzeitig zu einer sehr starken Filterung der Signale. Der Grund dafür ist die sehr unterschiedliche Energieverteilung in den einzelnen Frequenzbändern eines typischen Sprachsignals, wie in Abbildung 5.11 (a) dargestellt. Diese führt dazu, dass selbst bei einer gleich guten Trennung in allen Frequenzbändern die Bänder mit einer hohen Energie stark gedämpft werden.

Um dieses Problem zu umgehen, werden nun zwei weitere Vorschläge für die Bestimmung der Skalierungskoeffizienten gemacht. So können die Übersprechfaktoren

$$c'_i(\omega_k) = \tilde{S}_i(\omega_k) \quad (5.42)$$

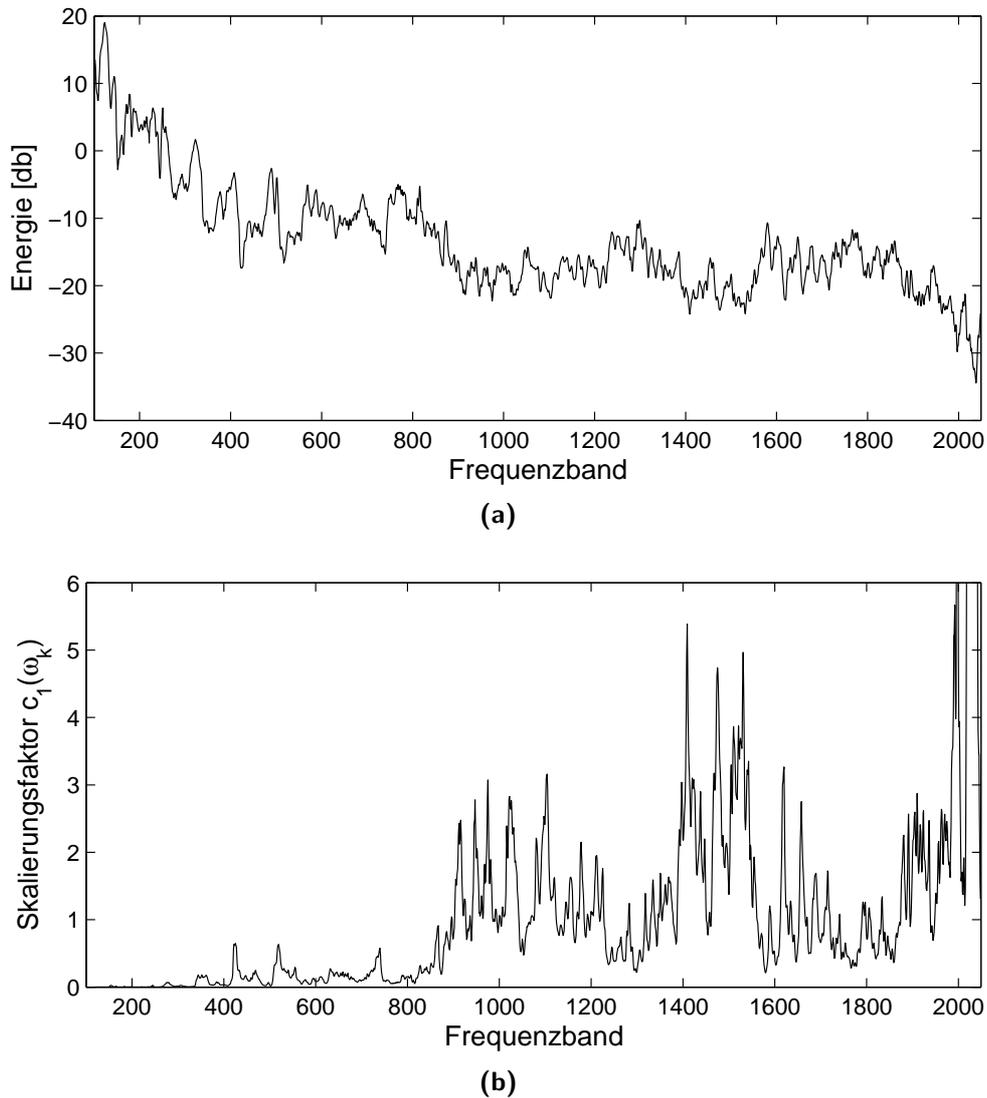


Abbildung 5.11: Beispiel für die Berechnung der Skalierungskoeffizienten mit der Max-SIR Methode. (a) Die Energieverteilung in den Frequenzbändern eines Sprachsignals. (b) Die resultierenden Skalierungskoeffizienten $c_1(\omega_k)$ nach Gleichung (5.40). Die tiefen Frequenzen werden unterdrückt, was zu einem hohen SIR aber auch starken Verzerrungen führt.

direkt als Skalierungskoeffizienten benutzt werden. Diese Vorgehensweise verbessert das SIR zwar nicht so stark, jedoch sind auch die Verzerrungen deutlich kleiner. Eine zweite Möglichkeit mit der Wahl

$$c_i''(\omega_k) = \min\{\tilde{S}_i(\omega_k), 1\} \quad (5.43)$$

Tabelle 5.1: Vergleich der Trennleistungen in dB der Max-SIR Methode, sowie der durch das Gesamtsystem (SFM_{ges}) und Entmischungssystem (SFM_{ent}) verursachten Verzerrungen.

	Links	Rechts	Gesamt	SFM _{ges}	SFM _{ent}
MDP	16.2	16.5	16.4	0.61	0.32
$c_i(\omega_k)$ aus Gl. (5.40)	29.9	24.1	27.8	0.02	0.02
$c'_i(\omega_k) = \tilde{S}(\omega_k)$	21.0	22.1	21.5	0.49	0.62
$c''_i(\omega_k) = \min\{\tilde{S}(\omega_k), 1\}$	20.8	21.7	21.3	0.55	0.61

verstärkt keine Frequenzbänder. Lediglich die nicht so gut getrennten Bereiche werden abgeschwächt. Dies führt in den gut getrennten Bereichen zu einer Skalierung nach dem MDP, was die kleinsten Verzerrungen in den getrennten Signalen liefert.

Beispiel (Fortsetzung)

Die blind geschätzte Trennleistung des Beispiels aus Kapitel 5.3.1 wurde dazu benutzt, um Skalierungskoeffizienten zu berechnen. Die Ergebnisse sind in den Abbildungen 5.11 und 5.12 dargestellt. In Abbildung 5.11 (a) ist die Energieverteilung eines der getrennten Signale gezeigt. Für diese initiale Berechnung wurde das MDP benutzt. Die nach Gleichung (5.42) berechneten Koeffizienten sind in Abbildung 5.11 (b) dargestellt. Es wird deutlich, dass die tiefen Frequenzen, welche eine hohe Energie und gleichzeitig eine schlechte Trennleistung haben, erheblich abgeschwächt werden. Dies führt zu einem sehr hohen SIR, aber auch einer starken Färbung der getrennten Signale. In Tabelle 5.1 sind die Ergebnisse zusammengefasst. Deutlich sichtbar ist die Verbesserung der Trennleistung um über 11dB. Dafür ist das Gesamtsystem, mit einem durchschnittlichen SFM von 0.02, sehr stark verzerrend.

Die Ergebnisse der weniger invasiven Methoden aus Gleichung (5.42) und (5.43) sind in Abbildung 5.12 dargestellt. Hier werden die tiefen Frequenzen nicht so stark gedämpft, so dass die Verzerrung des Gesamtsystems mit einem durchschnittlichen SFM von 0.49 beziehungsweise 0.54 wesentlich kleiner und vergleichbar mit dem MDP ist. Trotzdem ist die Trennleistung um 5dB besser, so dass dies die beste Skalierung für das Skalierungsproblem ist. Wenn das Entmischungssystem isoliert betrachtet wird, so ist die Skalierung mit den neuen

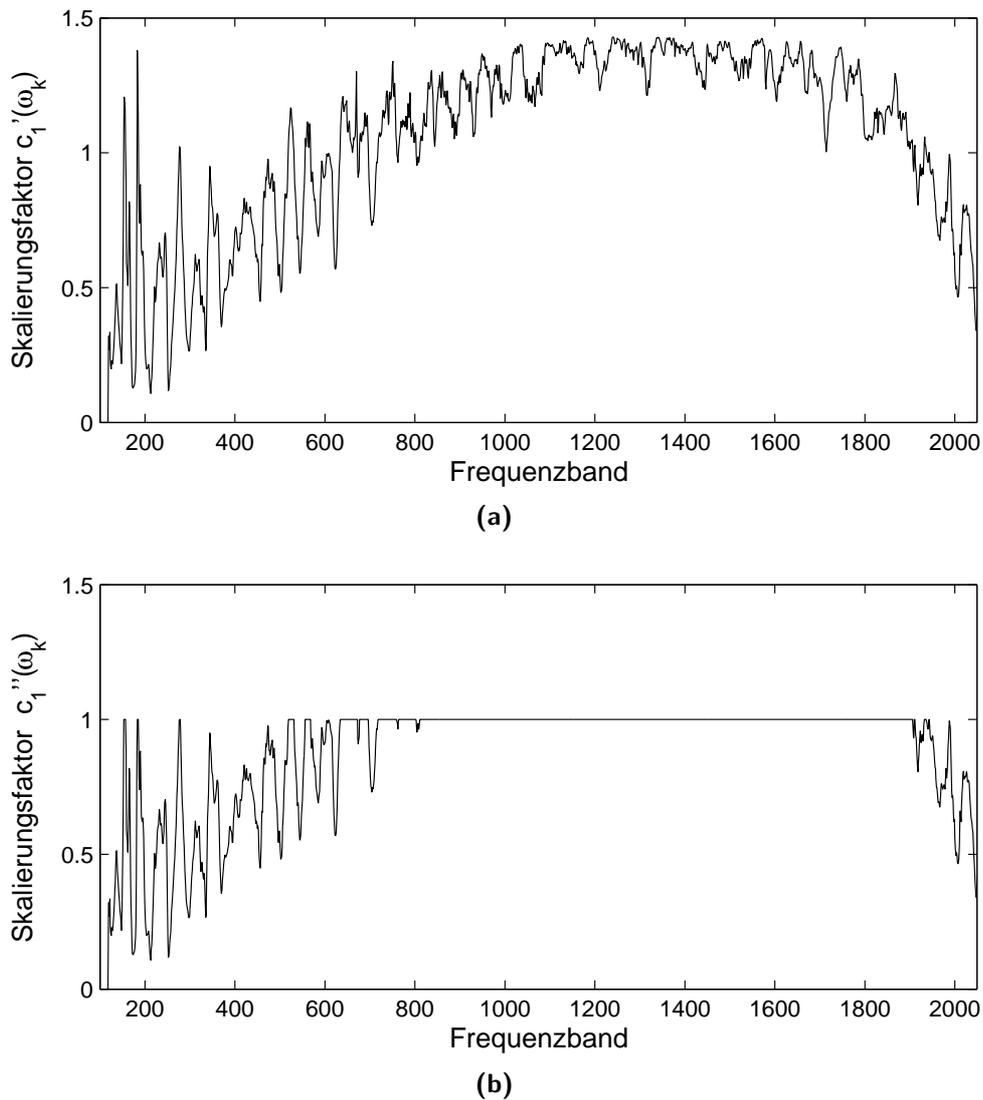


Abbildung 5.12: Beispiel für die Berechnung der Skalierungskoeffizienten mit der Max-SIR Methode. (a) Die Skalierungskoeffizienten nach Gleichung (5.42) und (b) nach Gleichung (5.43).

Methoden weniger verzerrend als die mit dem MDP.

Zusammenfassung

Die in diesem Abschnitt vorgestellte Methode zur blinden Schätzung der Trennleistung in den einzelnen Frequenzbändern eignet sich sehr gut, um die gesamte Trennleistung zu verbessern. Es wurden drei Methoden für die Berechnung der

Skalierungskoeffizienten vorgestellt. Neben der Möglichkeit der puren Maximierung des SIR, wurde insbesondere auch eine Methode vorgestellt, die das SIR deutlich verbessert und gleichzeitig die dadurch verursachten Verzerrungen klein hält.

5.4 Filterverkürzung

Channel Shortening (Kanalverkürzung) ist eine Technik, die dazu dient, mit Hilfe eines vor- oder nachgeschalteten Filters $w(n)$ die effektive Länge eines Übertragungskanals $h(n)$ zu verkürzen. Je nach Anforderungen an das Gesamtsystem

$$g(n) = w(n) * h(n) = h(n) * w(n) \quad (5.44)$$

werden verschiedene Entwurfsmethoden für $w(n)$ verwendet. So kann man versuchen, die Signale, die durch $h(n)$ gefiltert worden sind, vollständig zu rekonstruieren [70]. Dies bedeutet im Prinzip, dass das inverse System mit gegebenenfalls einer Verzögerung zu finden ist. Eine weniger strenge Anforderung ist mit der Aufgabe gegeben, die effektive Länge von $g(n)$ unter eine bestimmte Länge zu kürzen [56, 45]. Eine im akustischen Bereich verwendete Methode ist es, die Halleffekte auf ein nicht mehr wahrnehmbares Niveau zu reduzieren [57]. Da die Länge des Gesamtsystems $g(n)$ dazu nicht notwendigerweise reduziert werden muss, nennt man diese Technik auch *channel shaping*. Bei allen diesen Ansätzen kann $h(n)$ entweder direkt mit Hilfe von Testsequenzen gemessen oder durch Beobachtung der Ein- und Ausgangssignale geschätzt werden.

In den folgenden Abschnitten wird gezeigt, wie diese Ansätze im Bereich der blinden Quellentrennung benutzt werden können. Die beschriebenen Verfahren wurden bereits in [53] und [51] veröffentlicht. Hierbei wird jedoch versucht, die Entmischungsfiler direkt zu formen, ohne ein zusätzliches nachgeschaltetes System zu entwerfen (*filter shortening*, Filterverkürzung). Hierzu können die Skalierungskoeffizienten benutzt werden. Die Voraussetzung ist die vorherige Lösung des Permutationsproblems.

Im Weiteren wird gezeigt, wie der Zusammenhang zwischen den Skalierungskoeffizienten und den Entmischungsfilern im Zeitbereich hergestellt werden kann. Dieser Zusammenhang wird dann dazu benutzt, ein Kriterium zu definieren, mit dessen Hilfe die Skalierungskoeffizienten so gewählt werden können, dass die Entmischungsfiler möglichst kurz werden.

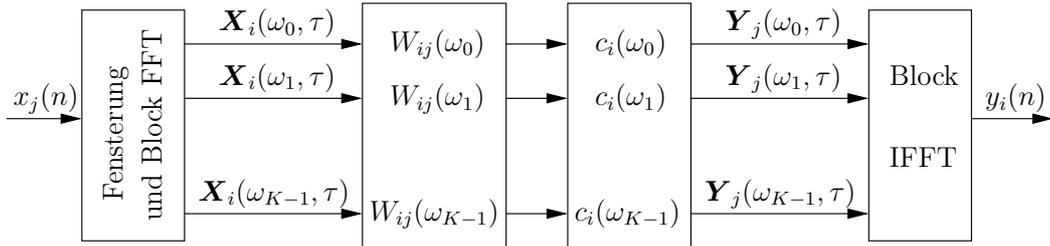


Abbildung 5.13: Ein Kanal des Entmischungssystems von der j -ten Mischung zu dem i -ten Ausgang, welches durch das Filter $w_{ij}(n)$ realisiert wird. Das Permutationsproblem wird in dieser Darstellung vernachlässigt.

5.4.1 Entmischungsfiler im Zeitbereich

Die bisher vorgestellten Methoden führen die Trennung im Zeit-Frequenzbereich durch, ohne die Entmischungsfiler im Zeitbereich explizit bestimmen zu müssen. Für die Filterverkürzung ist eine besondere Darstellung notwendig, bei der die Filterkoeffizienten $w_{ij}(n)$ im Zeitbereich in Abhängigkeit der Skalierungsfaktoren im Frequenzbereich $c_j(\omega_k)$ berechnet werden können. Dies kann erreicht werden, indem die Impulsantwort des Entmischungssystems bestimmt wird.

Das Entmischungssystem aus Kapitel 4 kann in einzelne Kanäle aufgespalten werden, welche durch die Filter $w_{ij}(n)$, wie in den Abbildungen 4.8 und 4.9 dargestellt, beschrieben werden. Die Koeffizienten $W_{ij}(\omega_k)$ dieser Filter werden mit Hilfe der ICA im Frequenzbereich bestimmt. Um die Impulsantworten $w_{ij}(n)$ zu bestimmen, reicht es nicht aus, $W_{ij}(\omega_k)$ mit Hilfe der IFFT in den Zeitbereich zu transformieren. Vielmehr muss der gesamte Kanal untersucht werden, da nur so die Wirkung der blockweisen Verarbeitung und damit auch die zirkulären Faltungsartefakte erfasst werden.

In Abbildung 5.13 ist so ein Kanal von der j -ten Mischung zu dem i -ten Ausgang dargestellt. Bei dieser Abbildung sind die Entmischungskoeffizienten $W_{ij}(\omega_k)$ bereits bestimmt, und auch das Permutationsproblem ist gelöst. Die einzigen unbekanntenen Elemente sind die Skalierungsfaktoren $c_i(\omega_k)$.

Alle Elemente des Entmischungssystems können mit Hilfe von Matrizen beschrieben werden. Mit

$$\mathbf{c}_i = [c_i(0), c_i(1), \dots, c_i(\omega_{K-1})]^T \quad (5.45)$$

5 Das Skalierungsproblem

und dem Operator $\text{diag}(\cdot)$, der einen Vektor in eine Diagonalmatrix und umgekehrt wandelt, können die Entmischungsfiler mit

$$\begin{aligned}
 \mathbf{w}_{ij} &= \sum_{l=0}^{L_\gamma/L_V} \Theta_l \cdot \mathcal{F}^{-1} \cdot \mathbf{C}_i \cdot \mathbf{W}_{ij} \cdot \mathcal{F} \cdot \Gamma_l \cdot \boldsymbol{\delta} \\
 &= \sum_{l=0}^{L_\gamma/L_V} \Theta_l \cdot \mathcal{F}^{-1} \cdot \text{diag}(\mathcal{F} \cdot \Gamma_l \cdot \boldsymbol{\delta}) \cdot \mathbf{W}_{ij} \cdot \mathbf{c}_i \\
 &= \mathbf{V}_{ij} \cdot \mathbf{c}_i
 \end{aligned} \tag{5.46}$$

berechnet werden. Für die Berechnung wurde zugrundegelegt, dass für diagonale Matrizen,

$$\mathbf{C} = \text{diag}(\mathbf{c}) \quad \text{und} \quad \mathbf{D} = \text{diag}(\mathbf{d}), \tag{5.47}$$

das Kommutativgesetz

$$\mathbf{CD} = \mathbf{DC} \tag{5.48}$$

gilt. Weiterhin wurde die Kommutativität der Multiplikation eines Vektors

$$\mathbf{c} = \text{diag}(\mathbf{C}) \tag{5.49}$$

mit einer Diagonalmatrix \mathbf{D} in der Form

$$\mathbf{Dc} = \mathbf{Cd} \tag{5.50}$$

benutzt. Die weiteren Elemente in der Gleichung (5.46) haben die folgende Bedeutung:

- $\boldsymbol{\delta}$ dient als Dirac-Impuls zur Messung der Impulsantwort und ist ein Vektor aus L_{FFT} Nullen und genau einer Eins an der Position L_γ .

$$\boldsymbol{\delta} = \begin{bmatrix} \mathbf{0}_{L_\gamma-1} \\ 1 \\ \mathbf{0}_{L_{\text{FFT}}-L_\gamma} \end{bmatrix} \tag{5.51}$$

Diese Wahl der Position stellt sicher, dass der Impuls alle bei der STFT relevanten Fenster

$$\boldsymbol{\gamma} = [\gamma(0), \gamma(1), \dots, \gamma(L_\gamma)]^T \tag{5.52}$$

trifft.

- Γ_l ist eine Diagonalmatrix mit den Koeffizienten des Fensters $\boldsymbol{\gamma}$, die beim l -ten Summanden um $l \cdot L_V$ auf der Hauptdiagonalen verschoben worden sind:

$$\Gamma_l = \text{diag} \left(\begin{array}{c} \mathbf{0}_{l \cdot L_V} \\ \boldsymbol{\gamma} \\ \mathbf{0}_{L_{\text{FFT}}-l \cdot L_V-L_\gamma} \end{array} \right). \tag{5.53}$$

- \mathcal{F} ist die Fouriermatrix und \mathcal{F}^{-1} die inverse Fouriermatrix. Bei reellen Signalen kann hier die Symmetrie ausgenutzt und \mathcal{F} entsprechend beschnitten werden. Die Größe von \mathcal{F} und \mathcal{F}^{-1} ist dann $K \times L_{\text{FFT}}$ beziehungsweise $L_{\text{FFT}} \times K$.
- \mathbf{W}_{ij} ist eine Diagonalmatrix mit den Frequenzkoeffizienten $W_{ij}(\omega_k)$ des ij -ten Entmischungsfilters:

$$\mathbf{W}_{ij} = \text{diag} \begin{pmatrix} W_{ij}(0) \\ W_{ij}(1) \\ \vdots \\ W_{ij}(\omega_k) \end{pmatrix}. \quad (5.54)$$

- \mathbf{C}_i ist eine Diagonalmatrix mit den Skalierungskoeffizienten des i -ten Ausgangs:

$$\mathbf{C}_i = \text{diag}(\mathbf{c}_i). \quad (5.55)$$

- Θ_l ist das Gegenstück zu Γ_l und dient dazu, die einzelnen Blöcke gegeneinander so zu verschieben, dass bei der Addition die korrekte Überlappung entsteht. Dies wird erreicht, indem die Identitätsmatrix \mathbf{I} mit der Größe $L_{\text{FFT}} \times L_{\text{FFT}}$ von oben und unten um eine der Blockverschiebung L_V entsprechende Anzahl von Nullen ergänzt wird:

$$\Theta_l = \begin{bmatrix} \mathbf{0}_{l \cdot L_V, L_{\text{FFT}}} \\ \mathbf{I}_{L_{\text{FFT}}, L_{\text{FFT}}} \\ \mathbf{0}_{L_{\text{FFT}} - l \cdot L_V, L_{\text{FFT}}} \end{bmatrix}. \quad (5.56)$$

Die Größe von Θ_l ist $L_{\text{FFT}} + L_{\text{V}} \times K$.

\mathbf{V}_{ij} fasst diese Elemente zu einer Matrix zusammen und ist eine Abbildung zwischen den frequenzbasierten Skalierungskoeffizienten $c_i(\omega_k)$ und den Entmischungsfilttern $w_{ij}(n)$ im Zeitbereich. \mathbf{V}_{ij} hat die gleiche Größe wie Θ_l .

5.4.2 Filter-Optimierung

Mit Hilfe der Matrix \mathbf{V}_{ij} können die Zeitbereichskoeffizienten der Entmischungsfiltter $w_{ij}(n)$ in Abhängigkeit von den Skalierungskoeffizienten $c_i(\omega_k)$ berechnet werden, was eine Optimierung dieser Filter erlaubt. Das Ziel der Filterverkürzung ist es, die Entmischungsfiltter so zu verändern, dass diese die Form eines Dirac-Impulses

$$\widetilde{\mathbf{w}}_{ij} = [0, \dots, 0, 1, 0, \dots, 0]^T \quad (5.57)$$

5 Das Skalierungsproblem

bekommen. Um die dazugehörige Skalierungskoeffizienten zu bestimmen, muss die Gleichung

$$\tilde{\mathbf{w}}_{ij} = \mathbf{V}_{ij} \mathbf{c}_i \quad (5.58)$$

nach \mathbf{c}_i gelöst werden. Die Gleichung (5.58) besitzt im Allgemeinen keine eindeutige Lösung, da \mathbf{V}_{ij} mehr Zeilen als Spalten hat. Ein weiteres Problem ergibt sich aus der Tatsache, dass die Skalierungskoeffizienten $c_i(\omega_k)$ auf alle Filter $W_{ij}(\omega_k)$, die zum Ausgang i führen, angewandt werden müssen. Daher ist eine gemeinsame Optimierung dieser Filter notwendig. Mit der Konkatenation der gewünschten Filter, die zum i -ten Ausgang führen, in der Form

$$\bar{\mathbf{w}}_i = \begin{bmatrix} \tilde{\mathbf{w}}_{i1} \\ \tilde{\mathbf{w}}_{i2} \\ \vdots \\ \tilde{\mathbf{w}}_{iN} \end{bmatrix}, \quad (5.59)$$

sowie der Konkatenation der dazugehörigen Matrizen \mathbf{V}_{ij} ,

$$\bar{\mathbf{V}}_i = \begin{bmatrix} \mathbf{V}_{i1} \\ \mathbf{V}_{i2} \\ \vdots \\ \mathbf{V}_{iN} \end{bmatrix}, \quad (5.60)$$

können die Skalierungskoeffizienten mit

$$\bar{\mathbf{w}}_i = \bar{\mathbf{V}}_i \mathbf{c}_i \quad (5.61)$$

bestimmt werden. Diese Gleichung ist nicht exakt lösbar, kann aber im Sinne des kleinsten quadratischen Fehlers durch die Minimierung von

$$\|\bar{\mathbf{w}}_i - \bar{\mathbf{V}}_i \mathbf{c}_i\|_{\ell_2} \quad (5.62)$$

berechnet werden. Dies kann zum Beispiel mit Hilfe der Pseudoinversen $\bar{\mathbf{V}}_i^+$ mit

$$\mathbf{c}_i = \bar{\mathbf{V}}_i^+ \bar{\mathbf{w}}_i \quad (5.63)$$

durchgeführt werden.

Die Position der Eins in den gewünschten Filtern $\tilde{\mathbf{w}}_{ij}$ ist nicht beliebig. Um ein optimales Ergebnis zu erzielen, sollte die Eins an der Position sein, an der die Entmischungsfiler \mathbf{w}_{ij} mit der Skalierung nach dem MDP das Maximum haben. Das Beibehalten der Position des Maximums führt zum Erhalt der Phase der Entmischungsfiler und ist somit besonders geeignet, die gewünschte Form zu erhalten.

Tabelle 5.2: Vergleich der Trennleistungen in dB mit der Filterverkürzungsmethode, sowie der durch das Gesamtsystem (SFM_{ges}) und Entmischungssystem (SFM_{ent}) verursachten Verzerrungen.

	Links	Rechts	Gesamt	SFM _{ges}	SFM _{ent}
MDP	16.2	16.5	16.4	0.61	0.32
Filter Shortening	22.1	26.3	24.2	0.33	0.65

5.4.3 Beispiel

Die Daten des letzten Abschnitts wurden zum zweiten Mal benutzt, um die Skalierungsfaktoren mit Hilfe der Filterverkürzungsmethode zu bestimmen. In den Abbildungen 5.14 und 5.15 sind die beiden benutzten Ansätze gegenübergestellt. In Abbildung 5.14 (a) sind die vier Entmischungsfiler eines (2×2)-Systems im Zeitbereich gezeigt. Es sind jeweils deutlich neben dem Hauptmaximum viele weitere große Koeffizienten zu erkennen. In der Abbildung 5.14 (b) sind die gleichen Filter mit der Skalierung nach der Filter Verkürzungsmethode dargestellt. Hier ist das Hauptmaximum deutlich prominenter, und die Energie der weiteren Koeffizienten ist deutlich kleiner. Dies kann auch sehr gut in Abbildung 5.15 beobachtet werden. Hier ist die Energie der Koeffizienten der Entmischungsfiler in Dezibel aufgetragen. Das Hauptmaximum ist nach wie vor sehr gut sichtbar. In dieser logarithmischen Darstellung ist jedoch die Reduktion der Energie in dem weit vom Hauptmaximum entfernten Koeffizienten deutlich. So konnte die Energie in dem hinteren Bereich der Impulsantworten im Schnitt um circa 15dB reduziert werden. Diese Reduktion hat insbesondere zur Folge, dass die zirkulären Artefakte gleichfalls um diese Größe reduziert werden.

Die Skalierungskoeffizienten des ersten Kanals sind in Abbildung 5.16 gezeigt. Es ist eine große Ähnlichkeit zu den Skalierungskoeffizienten mit der zweiten Variante der Max-SIR-Methode erkennbar. Allerdings werden hier die tiefen Frequenzen etwas stärker unterdrückt, was eine höhere Trennleistung zur Folge hat. Die Unterdrückung ist aber nicht so stark wie bei der ersten Variante der Max-SIR-Methode. Daraus folgt, wie Tabelle 5.2 darstellt, dass die Verzerrung an dem Gesamtsystems mit einem SFM-Wert von 0.33 nicht so stark ausfällt. Die Verzerrungen des Entmischungssystems sind ungefähr gleich groß.

5 Das Skalierungsproblem

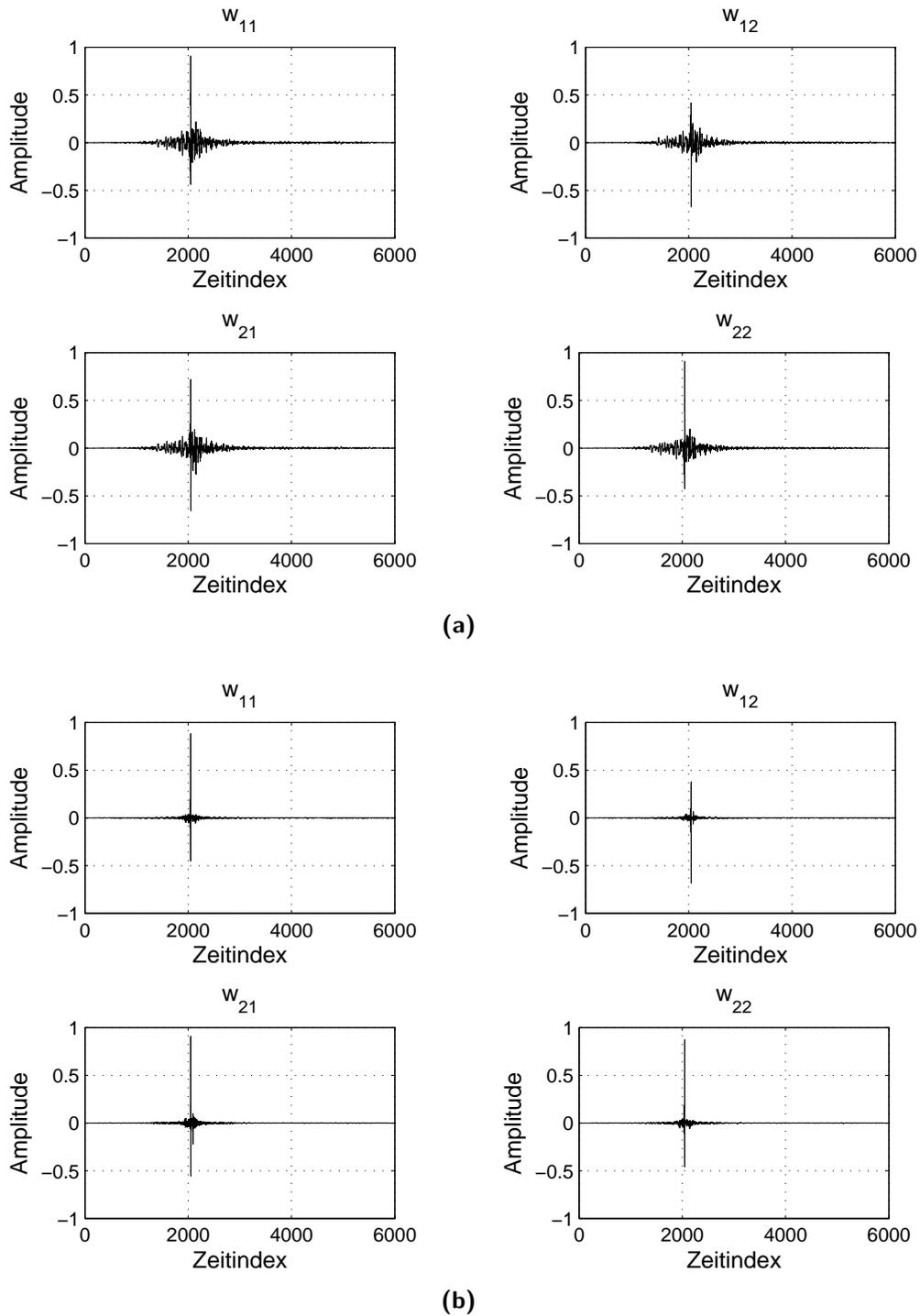


Abbildung 5.14: Die Verkürzung der Impulsantworten. (a) Die Impulsantworten berechnet mit Hilfe des MDP. (b) Filter nach der Filterverkürzungsmethode.

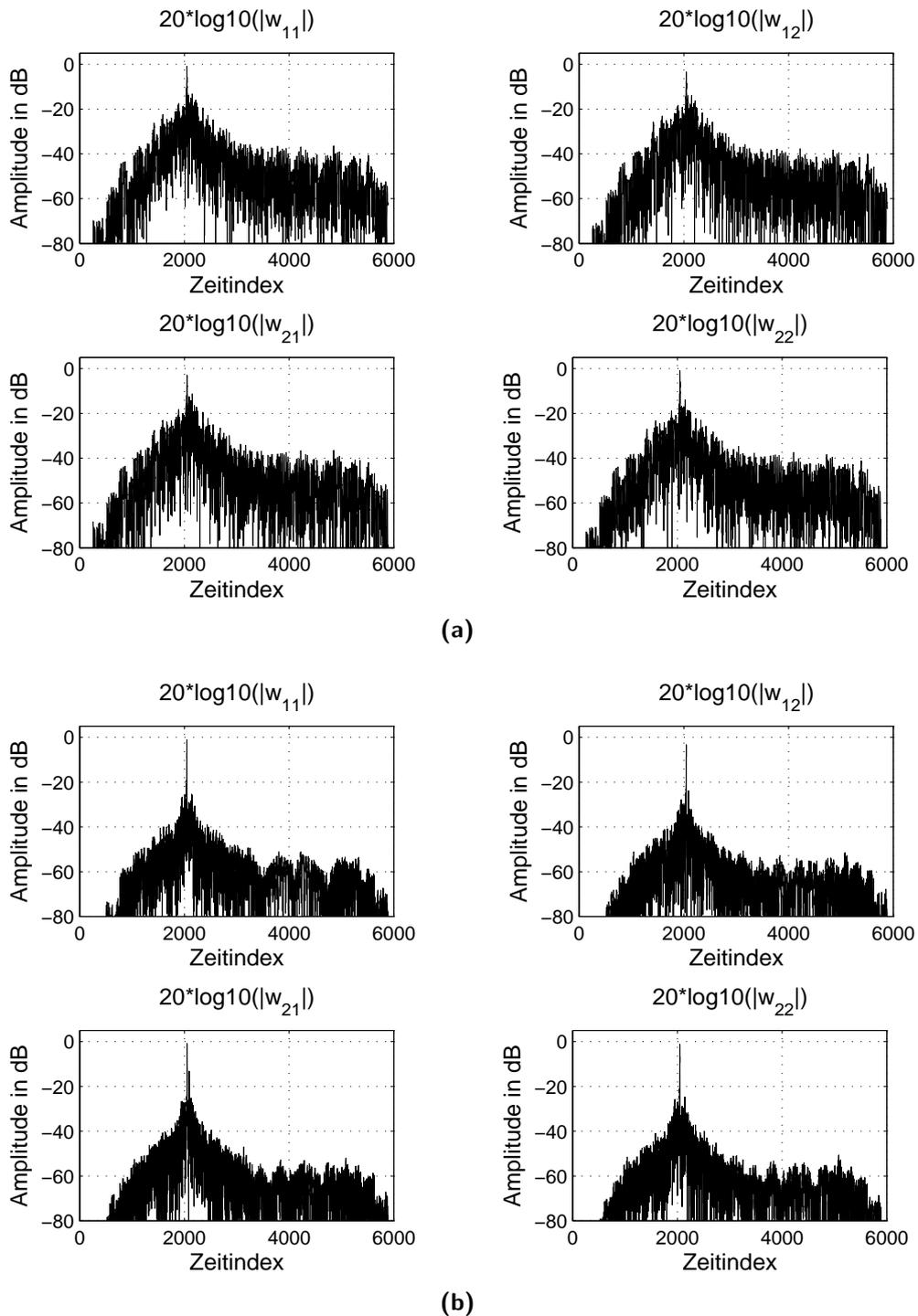


Abbildung 5.15: Alternative Darstellung der Filter aus Abbildung 5.14. Gezeigt sind die Energien der einzelnen Koeffizienten im Zeitbereich. (a) Die Impulsantworten berechnet mit Hilfe des MDP. (b) Filter nach der Filterverkürzungsmethode.

5 Das Skalierungsproblem

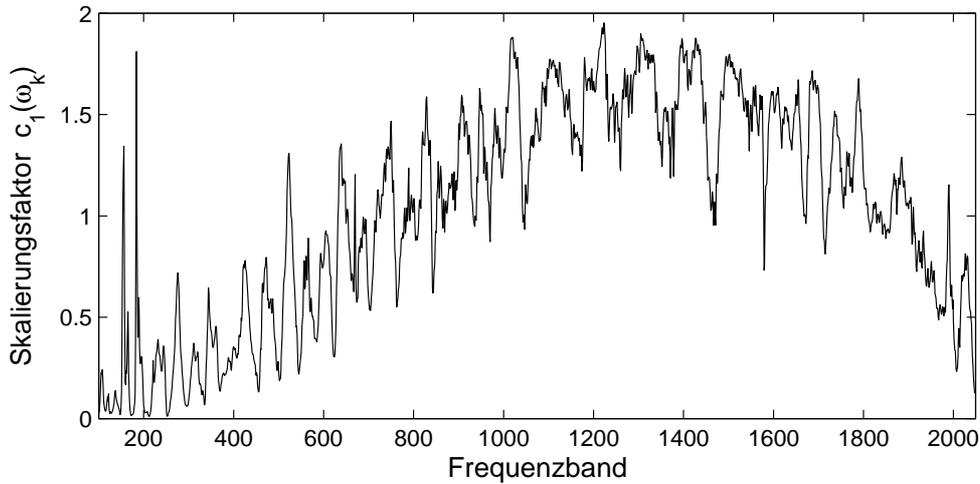


Abbildung 5.16: Beispiel für die Berechnung der Skalierungsfaktoren mit Hilfe der Filterverkürzungsmethode.

Zusammenfassung

In diesem Abschnitt wurde der Zusammenhang zwischen den Skalierungskoeffizienten im Frequenzbereich und den Zeitbereichskoeffizienten der Entmischungsfiler hergeleitet. Dieser Zusammenhang wurde dazu benutzt, die Entmischungsfiler in die Form eines Dirac-Impulses zu bringen. An einem Beispiel wurde gezeigt, dass die Methode in der Lage ist, die ungewollten Koeffizienten der Entmischungsfiler um ungefähr 15dB zu verkleinern.

5.5 Filterformung

Im letzten Abschnitt wurden die Impulsantworten so verändert, dass das Hauptmaximum möglichst groß und alle anderen Koeffizienten gleichzeitig möglichst klein werden. Bei der Minimierung von (5.62) werden alle unerwünschten Koeffizienten gleich behandelt. Eine alternative Methode, bei der die einzelnen Koeffizienten eine unterschiedliche Gewichtung bekommen, ist das *filter shaping* (Filterformung). Es basiert auf den Ideen des *channel shaping* (Kanalformung) [8, 38, 57]. In dem folgenden Abschnitt wird gezeigt, wie diese Methode im Bereich der blinden Quellentrennung benutzt werden kann. Das Verfahren wurde bereits in [51] beschrieben.

5.5.1 Optimierung der Entmischungsfiler

Bei der klassischen Variante der Kanalformung wird mit Hilfe eines Fensters

$$\boldsymbol{\gamma}_d = [\gamma_d(0), \gamma_d(1), \dots, \gamma_d(L_W - 1)]^T \quad (5.64)$$

ein erwünschter (*desired*) Bereich der Impulsantwort definiert, in dem die Impulsantwort beliebige Werte annehmen kann. Parallel dazu wird mit

$$\boldsymbol{\gamma}_u = [\gamma_u(0), \gamma_u(1), \dots, \gamma_u(L_W - 1)]^T \quad (5.65)$$

der unerwünschte (*undesired*) Bereich definiert, in dem die Impulsantwort möglichst kleine Werte haben soll. Durch verschiedene Formen der Fenster $\boldsymbol{\gamma}_d$ und $\boldsymbol{\gamma}_u$ können die einzelnen Bereiche der Impulsantwort unterschiedlich stark gewichtet werden. Das Ziel der Filterformung ist es, die Energie im erwünschten Bereich zu maximieren und gleichzeitig die Energie im unerwünschten Bereich zu minimieren.

Mittels der Matrix \mathbf{V}_{ij} aus dem letzten Abschnitt, die den Zusammenhang zwischen den Skalierungskoeffizienten \mathbf{c}_i und der Impulsantwort \mathbf{w}_{ij} herstellt, kann der erwünschte Teil der Impulsantwort mit

$$\mathbf{w}_{d_{ij}} = \text{diag}(\boldsymbol{\gamma}_{d_{ij}}) \mathbf{V}_{ij} \mathbf{c}_i \quad (5.66)$$

und der unerwünschte Teil mit

$$\mathbf{w}_{u_{ij}} = \text{diag}(\boldsymbol{\gamma}_{u_{ij}}) \mathbf{V}_{ij} \mathbf{c}_i \quad (5.67)$$

berechnet werden. Hierbei können der erwünschte und unerwünschte Bereich für jede Impulsantwort \mathbf{w}_{ij} einzeln mit $\boldsymbol{\gamma}_{d_{ij}}$ und $\boldsymbol{\gamma}_{u_{ij}}$ definiert werden. Um eine gemeinsame Optimierung der Filter \mathbf{w}_{ij} , die zu einem Ausgang führen, zu erreichen, wird wie im letzten Abschnitt die Konkatenation der Elemente benutzt. Hierzu wird $\bar{\mathbf{V}}_i$ wie in Gleichung (5.60) definiert. Zudem werden die Vektoren

$$\bar{\boldsymbol{\gamma}}_{d_i} = \begin{bmatrix} \boldsymbol{\gamma}_{d_{i1}} \\ \boldsymbol{\gamma}_{d_{i2}} \\ \vdots \\ \boldsymbol{\gamma}_{d_{iN}} \end{bmatrix} \quad (5.68)$$

sowie

$$\bar{\mathbf{w}}_{d_i} = \begin{bmatrix} \mathbf{w}_{d_{i2}} \\ \mathbf{w}_{d_{i3}} \\ \vdots \\ \mathbf{w}_{d_{iN}} \end{bmatrix} \quad (5.69)$$

5 Das Skalierungsproblem

eingeführt. Analog wird mit $\bar{\gamma}_{u_i}$ und $\bar{\omega}_{u_i}$ verfahren. Die Energie in dem erwünschten Bereich kann als

$$\begin{aligned}\bar{\omega}_{d_i}^H \bar{\omega}_{d_i} &= \mathbf{c}_i^H \bar{\mathbf{V}}_i^H \text{diag}(\bar{\gamma}_{d_i}^H) \text{diag}(\bar{\gamma}_{d_i}) \bar{\mathbf{V}}_i \mathbf{c}_i \\ &= \mathbf{c}_i^H \bar{\Psi}_i \mathbf{c}_i\end{aligned}\quad (5.70)$$

und in dem unerwünschten Bereich mit

$$\begin{aligned}\bar{\omega}_{u_i}^H \bar{\omega}_{u_i} &= \mathbf{c}_i^H \bar{\mathbf{V}}_i^H \text{diag}(\bar{\gamma}_{u_i}^H) \text{diag}(\bar{\gamma}_{u_i}) \bar{\mathbf{V}}_i \mathbf{c}_i \\ &= \mathbf{c}_i^H \bar{\Phi}_i \mathbf{c}_i\end{aligned}\quad (5.71)$$

berechnet werden. Die Maximierung von $\mathbf{c}_i^H \bar{\Psi}_i \mathbf{c}_i$ bei gleichzeitiger Minimierung von $\mathbf{c}_i^H \bar{\Phi}_i \mathbf{c}_i$ führt zu dem folgenden generalisierten Eigenwertproblem [56]:

$$\bar{\Phi}_i \cdot \mathbf{c}_i^{opt} = \bar{\Psi}_i \cdot \mathbf{c}_i^{opt} \cdot \lambda_{max}. \quad (5.72)$$

Der Vektor mit den optimalen Skalierungskoeffizienten \mathbf{c}_i^{opt} ist der Eigenvektor, der zu dem größten Eigenwert λ_{max} gehört.

5.5.2 Fensterfunktionen

Die Form der Impulsantwort kann mit Hilfe des erwünschten Fensters γ_d in Verbindung mit dem unerwünschten Fenster γ_u bestimmt werden. Es können unterschiedliche Formen gewählt werden, je nach dem Zweck, für den die Filter benutzt werden.

Maximierung des Betrages des Hauptmaximums

Die einfachste Fensterform ist in Abbildung 5.17 dargestellt. Das Fenster γ_d , welches den erwünschten Bereich auswählt, ist ein Dirac-Impuls, dessen Position die Position des Maximums, des nach dem MDP bestimmten Entmischungsfilters, ist. Das Fenster γ_u , welches den unerwünschten Bereich auswählt, ist das Komplement dazu:

$$\gamma_u = 1 - \gamma_d \quad (5.73)$$

Dieser Ansatz ist dem Filterverkürzungsmethode nachempfunden und liefert nahezu identische Ergebnisse.

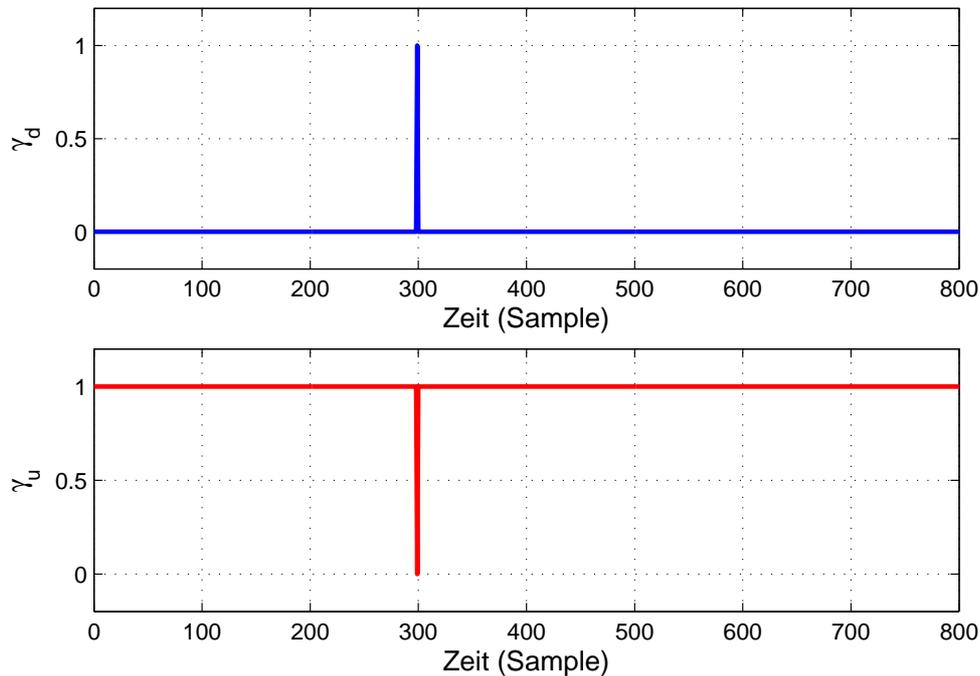


Abbildung 5.17: Fenster für die Maximierung des Betrages des Hauptmaximums. Oben das Fenster γ_d für den erwünschten Bereich und unten das komplementäre Fenster γ_u , welches den unerwünschten Bereich auswählt.

Konzentration auf einen Bereich

Ein anderer Ansatz ist es, die Energie nicht nur im Hauptmaximum zu konzentrieren, sondern über einen kurzen Bereich zu verteilen [45]. Das Fenster zur Gewinnung des erwünschten Anteils der Impulsantwort ist ein Rechteck der Länge L_r gefolgt von L_p Nullen.

$$\gamma_d = \begin{bmatrix} \mathbf{1}_{L_r} \\ \mathbf{0}_{L_p} \end{bmatrix}. \quad (5.74)$$

γ_u ist weiterhin das Komplement zu γ_d , wie in Abbildung 5.18 dargestellt. Bei diesem Ansatz können die Filterkoeffizienten in dem gesamten erwünschten Bereich beliebig verteilt werden. Wegen der größeren Freiheit können die Koeffizienten so gewählt werden, dass die Unterdrückung außerhalb dieses Bereiches besser ist. Leider ist dieser Ansatz aufgrund der so erlaubten Echos nur bedingt für akustische Anwendungen geeignet [54].

5 Das Skalierungsproblem

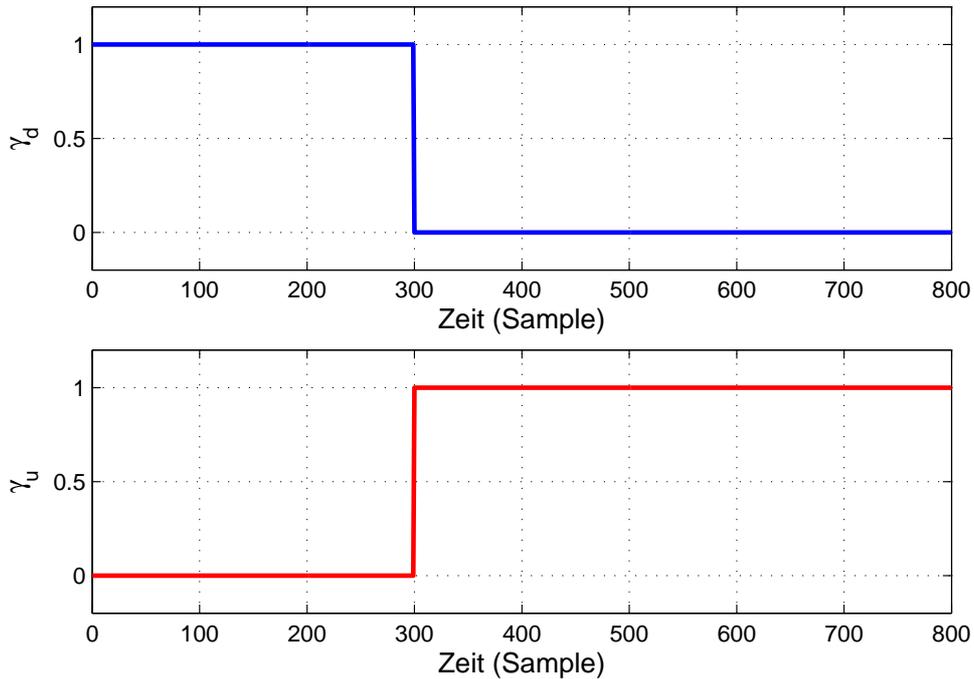


Abbildung 5.18: Fenster für die Konzentration der Energie in einem Bereich. Diese Wahl der Fenster erlaubt eine bessere Unterdrückung außerhalb des erwünschten Bereiches.

Exponentielle Verläufe

Im Gebiet der akustischen Kanalformung können exponentiell abfallende Fenster mit

$$\gamma_d(n) = \begin{cases} 10^{q_d(n_0-n)} & \text{für } 0 \leq n \leq n_0 \\ 0 & \text{sonst} \end{cases} \quad (5.75)$$

und

$$\gamma_u(n) = \begin{cases} 10^{q_u(n_0-n)} & \text{für } 0 \leq n \leq n_0 \\ 0 & \text{sonst} \end{cases} \quad (5.76)$$

benutzt werden [38, 57]. In Abbildung 5.19 sind diese Fenster mit $q_d = -0.005$ und $q_u = 0.0015$ dargestellt. Die Benutzung exponentiell abfallender Fenster hat das Ziel, die Impulsantwort in eine Form zu bringen, die einer echten Raumimpulsantwort nachempfunden ist. Das Fenster für den erwünschten Bereich hat unmittelbar diese Form. Darüber hinaus werden die weiter entfernten Koeffizienten durch das exponentielle Ansteigen des Fensters für den unerwünschten Teil stärker unterdrückt. Im Beispiel in Abbildung 5.19 hat die Impulsantwort

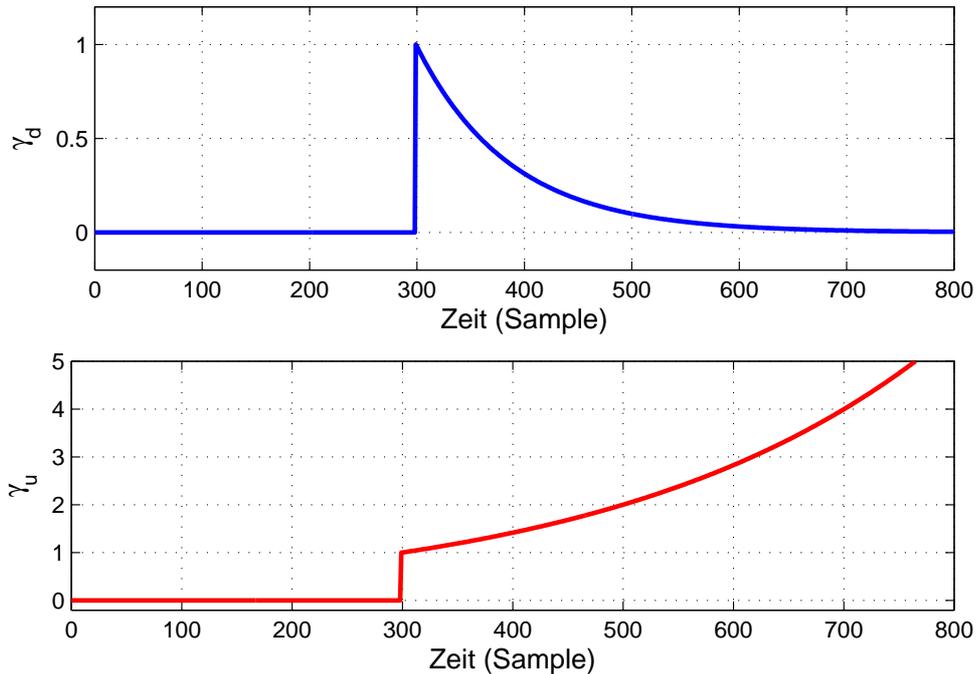


Abbildung 5.19: Fenster für Filterformung mit exponentiellen Verläufen. Diese Form führt zu Impulsantworten, die Raumimpulsantworten nachempfunden sind.

eine Verzögerung um 300 Koeffizienten. Der vorausgehende Bereich ist bei der Kanalformung irrelevant und kann für beide Fenster auch mit einer Folge von Einsen gefüllt werden [38].

Doppelt exponentielle Verläufe

Die Entmischungsfiler der blinden Quellentrennung haben üblicherweise auch vor dem Hauptmaximum viele Koeffizienten ungleich null, wie in Abbildung 5.20 (a) dargestellt. Deswegen wird dieser Bereich ebenfalls mit Hilfe von Exponentiellfunktionen geformt. Damit ergibt sich ein doppelt exponentieller Verlauf mit

$$\gamma_d(n) = \begin{cases} 10^{q_1(n_o-n)} & \text{für } 0 \leq n \leq n_0 \\ 10^{q_2(n-n_o)} & \text{für } n_0 \leq L_W \end{cases} \quad (5.77)$$

und

$$\gamma_u(n) = \begin{cases} 10^{q_3(n_o-n)} & \text{für } 0 \leq n \leq n_0 \\ 10^{q_4(n-n_o)} & \text{für } n_0 \leq L_W. \end{cases} \quad (5.78)$$

Tabelle 5.3: Vergleich der Trennleistungen in dB mit dem filter shaping Prinzip, sowie der durch das Gesamtsystem (SFM_{ges}) und Entmischungssystem (SFM_{ent}) verursachten Verzerrungen.

	Links	Rechts	Gesamt	SFM _{ges}	SFM _{ent}
MDP	16.2	16.5	16.4	0.61	0.32
Filter Shortening	22.1	26.3	24.2	0.33	0.65
Filter Shaping	25.0	28.8	26.8	0.21	0.47

In Abbildung 5.20 (b) sind diese Fenster mit $q_1 = -0.005$, $q_2 = 0.0005$, $q_3 = 0.01$ und $q_4 = -0.001$ dargestellt. Die unterschiedliche Gewichtung der Bereiche vor und nach dem Hauptmaximum führt dazu, dass der vordere Bereich stärker unterdrückt wird als der nachfolgende. Diese Form ist psychoakustisch motiviert. So führen die Koeffizienten vor dem Hauptmaximum zu den sogenannten Vor-Echos (pre-echoes), die deutlich störender empfunden werden als die Echos, die durch die dem Hauptmaximum folgenden Koeffizienten verursacht sind. Die Position n_0 des Hauptmaximums wird mit Hilfe des MDP bestimmt.

5.5.3 Beispiel

Die Daten aus dem Abschnitt 5.3 wurden ein drittes Mal benutzt, um die Skalierungsfaktoren mit Hilfe der Filterformung zu bestimmen. Die gewonnenen Filter sind in Abbildung 5.21 (a) dargestellt. In der Zeitdarstellung haben diese Filter eine große Ähnlichkeit zu den Filtern aus Abbildung 5.14, die mit Hilfe der Filterverkürzung berechnet worden sind. Auch hier ist das Hauptmaximum deutlich prominenter, und die anderen Koeffizienten werden deutlich verkleinert. Ein wesentlicher Unterschied ist jedoch in den Energien der vom Hauptmaximum weiter entfernten Koeffizienten sichtbar. Mit Hilfe der Filterformung konnten die unerwünschten Koeffizienten um weitere 10dB gesenkt werden, so dass insgesamt eine Reduktion der Energie um ca. 25dB erreicht werden konnte.

Die Skalierungskoeffizienten für den ersten Kanal sind in Abbildung 5.22 dargestellt. Die tiefen Frequenzen werden noch stärker als bei der Filterverkürzung unterdrückt. Daraus folgt eine etwas bessere Trennleistung. Aus den Ergebnissen aus Tabelle 5.3 geht hervor, dass die Trennleistung fast so gut wie nach dem Max-SIR Algorithmus ist. Jedoch sind die verursachten Störungen mit

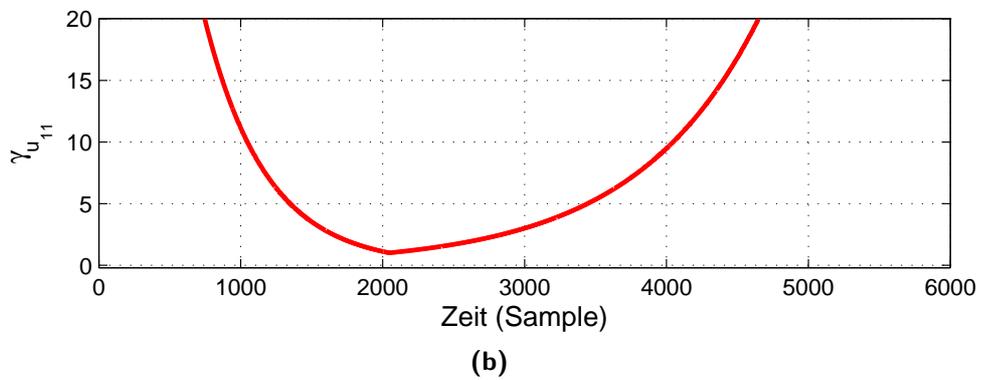
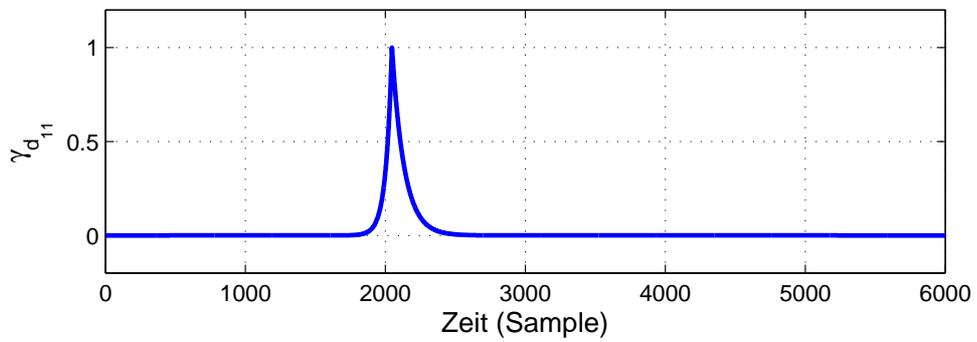
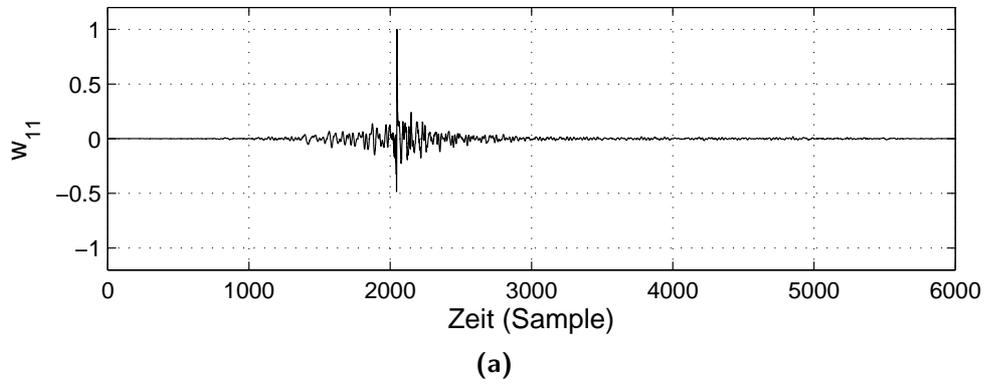


Abbildung 5.20: Doppelt exponentielle Fenster für die Filterformung. Unterschiedlichen Exponenten führen zu Impulsantworten, bei denen der Bereich vor dem Hauptmaximum stärker als der dahinter gedämpft wird. (a) Die zu formende Impulsantwort. (b) Die doppelt exponentiellen Verläufe der Fenster ausgerichtet auf das Hauptmaximum der Impulsantwort.

5 Das Skalierungsproblem

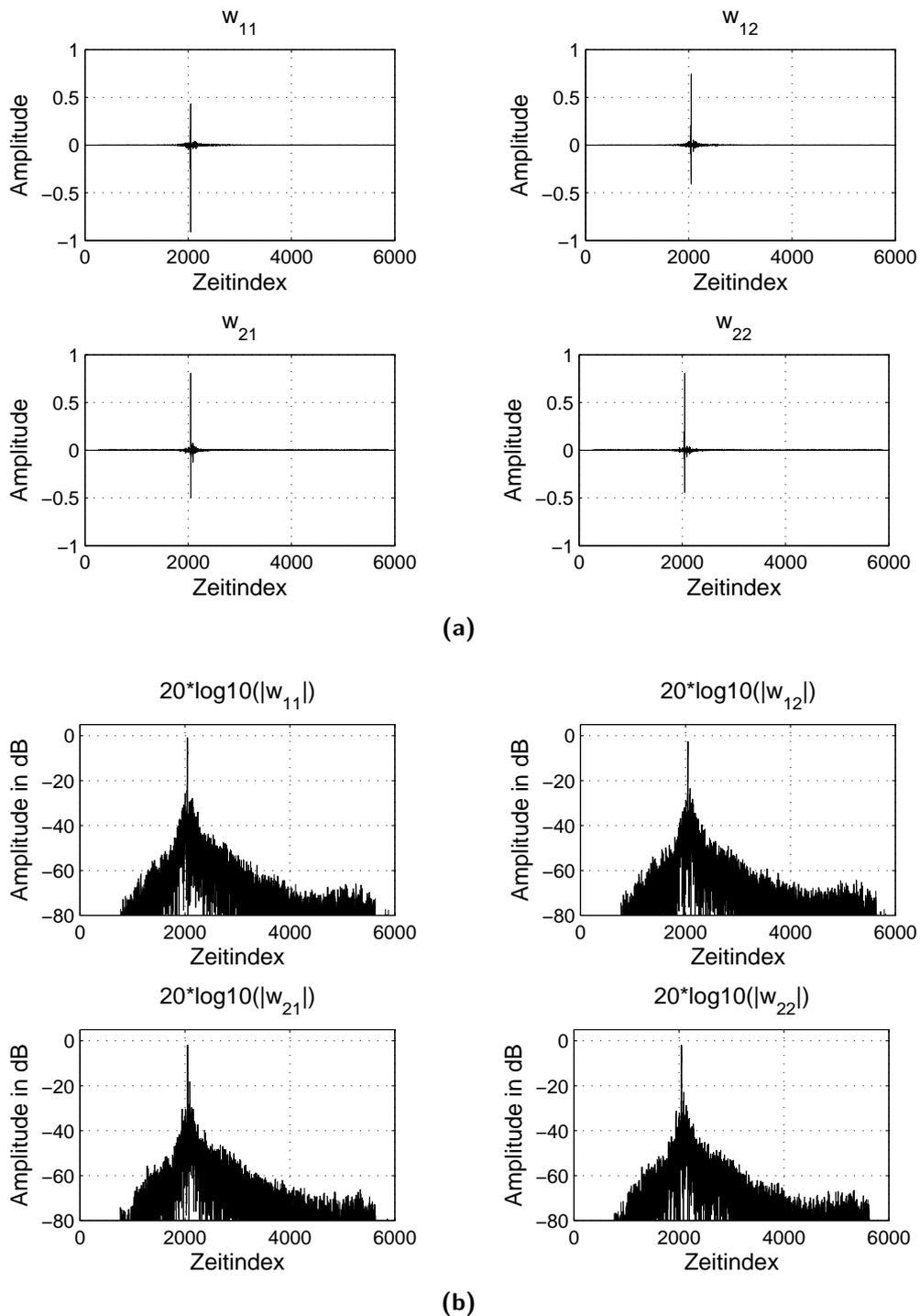


Abbildung 5.21: Beispiel für Filterformung. (a) Die Filter im Zeitbereich. (b) Die Energie der einzelnen Koeffizienten. Der Vergleich mit Abbildung 5.15 zeigt bessere Dämpfung der weiter entfernten Koeffizienten.

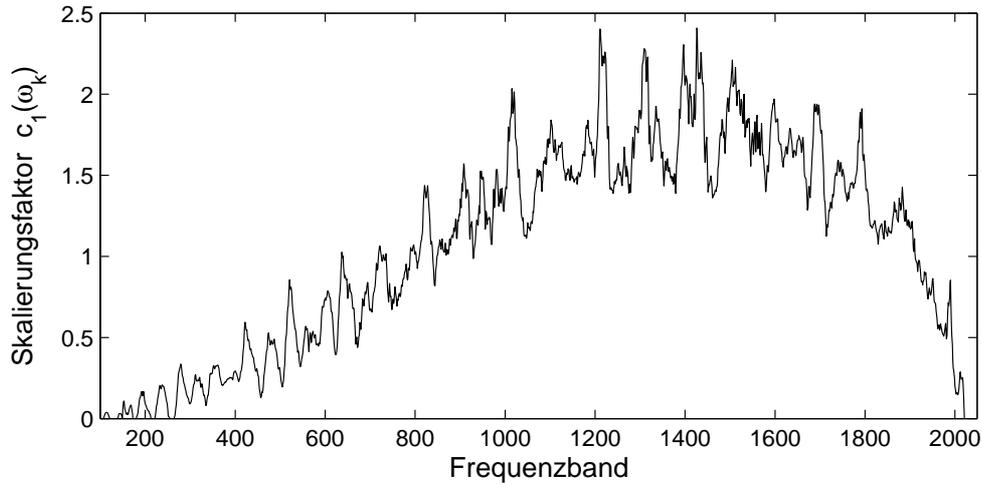


Abbildung 5.22: Beispiel für die Berechnung der Skalierungsfaktoren mit Hilfe der Filterformung.

einem SFM von 0.21 für das Gesamtsystem beziehungsweise 0.47 für das Entmischungssystem erheblich kleiner. Der hörbare Eindruck der getrennten Daten ist besser als diese Messwerte es wiedergeben. So sind die hörbaren Echos deutlich kleiner, und die Verzerrung des Frequenzganges wirkt daher vergleichsweise weniger störend. Für ein objektives Maß für die Echos, das den subjektiven Eindruck widerspiegelt, müssten psychoakustische Modelle benutzt werden, welche aber im Rahmen dieser Arbeit nicht weiter betrachtet werden.

5.6 Zusammenfassung

In diesem Kapitel wurden drei neue Methoden für die Lösung des Skalierungsproblems vorgestellt. Die erste Methode kann eingesetzt werden, ohne dass eine Lösung des Permutationsproblems nötig ist. Mit Hilfe einer blinden Schätzung der Trennleistung in jedem Frequenzband werden die Skalierungskoeffizienten dabei so bestimmt, dass die Gesamttrennleistung maximiert wird.

Die beiden anderen Methoden setzen die Lösung des Permutationsproblems voraus. Sie erlauben die Formung der Impulsantworten mit dem Ziel, das Hauptmaximum zu maximieren und gleichzeitig die anderen Koeffizienten zu minimieren. Die so gewonnenen Impulsantworten verbessern gleichzeitig die Trennleistung, haben aber keine so großen Verzerrungen wie die Max-SIR Methode zur Folge. Der Höreindruck wird verbessert.

5 *Das Skalierungsproblem*

Das Permutationsproblem

Das zweite zentrale Problem der konvolutiven blinden Quellentrennung ist das Permutationsproblem. Nach der Trennung jedes Frequenzbandes mit Hilfe der instantanen ICA ist die Reihenfolge der Signale in jedem Frequenzband zufällig. Vor der Transformation in den Zeitbereich muss diese Reihenfolge in allen Frequenzbändern aber gleich sein, denn anderenfalls würden mehrere Spektralanteile zu einem Zeitsignal zusammengefasst werden. Folglich wäre dieses Zeitsignal gemischt und das ganze Verfahren wäre fehlgeschlagen. Das äußere Permutationsproblem, bei dem die gesamten Signale vertauscht sein können, ist von dem angesprochenen inneren Permutationsproblem unabhängig. Das äußere Permutationsproblem kann mit blinden Methoden nicht gelöst werden und wird hier nicht weiter betrachtet.

N Quellen können in jedem Frequenzband auf $N!$ verschiedene Weisen permutiert sein. Bei K Frequenzbändern sind damit also $(N!)^K$ verschiedene Permutationen der Ausgangssignale möglich. In einem realistischen Szenario mit einer FFT-Länge von 8192 und damit 4097 Frequenzbändern existieren bei zwei Quellen also ungefähr $4 \cdot 10^{149}$ Möglichkeiten, die Frequenzbänder zu permutieren. Es ist nicht möglich, alle diese Permutationen durchzutesten. Eine Möglichkeit, dieses Problem zu umgehen, ist es, jeweils nur kleine Untermengen der Frequenzbänder zu ordnen und dann gruppenweise gegeneinander zu entpermutieren.

Die existierenden Ansätze zur Lösung des Permutationsproblems können in zwei Gruppen eingeteilt werden. Zur ersten Gruppe gehören Algorithmen, die die Eigenschaften der getrennten Signale benutzen. Ein Beispiel hierfür ist die Annahme der Korreliertheit benachbarter Frequenzbänder [32, 72]. Die zweite Gruppe der Algorithmen nutzt hingegen die Eigenschaften des Entmischungssystems. So kann beispielsweise die Ähnlichkeit der Entmischungsmatrizen der benachbarten Frequenzbänder ausgenutzt werden. In [73] wurden die Matrizen des

6 Das Permutationsproblem

Entmischungssysteme als *Beamformer* interpretiert, die bestimmte räumliche Richtungen unterdrücken. Dieses Prinzip wurde in [80, 34] zu den sogenannten *directivity patterns* erweitert, und mit Hilfe von Nah- und Fernfeldmodellen konnten in [61] sogar Quellen aus einer Richtung separiert werden. Diese Ansätze bedürfen jedoch der Kenntnis über die Positionen der Mikrofone zueinander, so dass es sich nicht um vollständig blinde Verfahren handelt. In [55] wurde ein Ansatz vorgestellt, welcher die Permutation vermeidet. Dies wird erreicht, indem die Entmischungsmatrizen nach jedem Iterationsschritt durch eine Projektion auf zeitlich beschränkte Filter zueinander in Verbindung gebracht werden. Der Nachteil dieses Verfahrens ist die oft starke Färbung der getrennten Signale [47].

In diesem Kapitel wird ein neuer Algorithmus der ersten Gruppe vorgestellt. Die grundsätzliche Überlegung ist, wie in [48] vorgestellt, die Modellierung jedes Frequenzbandes mit Hilfe der GGD. So wurde in [48] gezeigt, dass sich die Parameter der GGD in den benachbarten Frequenzbändern nur wenig unterscheiden, und daraus wird ein Entpermutationsalgorithmus auf Basis des β -Parameters hergeleitet. In [50] wurde der Ansatz für den Parameter α erweitert. In [52] wurde eine Vereinfachung vorgestellt, welche durch ein gemeinsames Kriterium für beide Parameter zu einer einfacheren Berechnung führt.

Problemdefinition

Die Lösung des inneren Permutationsproblems besteht darin, für jedes Frequenzband ω_k eine Permutationsmatrix $\mathbf{P}(\omega_k)$ der Größe $N \times N$ zu finden, so dass die Gesamttrennleistung nach Gleichung (4.42), nach nicht-blinder Lösung des äußeren Permutationsproblems, maximal wird.

Eine Permutationsmatrix ist quadratisch und hat in jeder Zeile und jeder Spalte genau eine Eins, und der Rest besteht aus Nullen. Es existieren genau $N!$ verschiedene Permutationsmatrizen. Sie können durch ein zeilenweises Vertauschen der Identitätsmatrix gewonnen werden. In dem (2×2) -Fall existieren zwei Permutationsmatrizen

$$\mathbf{P}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{und} \quad \mathbf{P}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (6.1)$$

In dem (3×3) -Fall wächst die Anzahl auf sechs Möglichkeiten:

$$\begin{aligned} \mathbf{P}_1 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, & \mathbf{P}_2 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, & \mathbf{P}_3 &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ \mathbf{P}_4 &= \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, & \mathbf{P}_5 &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, & \mathbf{P}_6 &= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \end{aligned} \quad (6.2)$$

Die Multiplikation einer Permutationsmatrix mit einer Entmischungsmatrix mit

$$\mathbf{W}^i(\omega_k) = \mathbf{P}_i(\omega_k) \mathbf{W}(\omega_k) \quad (6.3)$$

führt dazu, dass die Zeilen der Entmischungsmatrix nach der i -ten Permutation getauscht werden und damit auch die Reihenfolge der getrennten Signale $\mathbf{Y}(\omega_k, \tau)$ in den jeweiligen Frequenzbändern vertauscht wird.

6.1 Nicht-blinde Verfahren

Um die Leistung der Entpermutationsalgorithmen beurteilen zu können, werden zunächst zwei nicht-blinde Methoden vorgestellt, um die bestmögliche Permutation zu bestimmen. Mit Hilfe eines Vergleichs der beiden Verfahren wird im Weiteren noch gezeigt, dass keine eindeutig *beste* Permutation existiert.

6.1.1 Entpermutation nach der Trennleistung

Die erste nicht-blinde Methode basiert auf der Idee, dass die Maximierung der Trennleistung in jedem Frequenzband zu einer maximalen Gesamttrennleistung führt. Hierzu muss lediglich die Trennleistung

$$\text{SIR}(\omega_k) = \text{SIR}(\mathbf{W}(\omega_k), \mathbf{S}(\omega_k, \tau)) \quad (6.4)$$

in Abhängigkeit von der Entmischungsmatrix $\mathbf{W}(\omega_k)$ und den bekannten Quellsignalen $\mathbf{S}(\omega_k, \tau)$ für jedes Frequenzband nach Gleichung (4.42) für alle möglichen Permutationen berechnet werden. Die beste Permutation kann dann mit

$$\mathbf{P}_{\text{SIR}}(\omega_k) = \underset{\mathbf{P}_i}{\text{argmax}} \{ \text{SIR}(\mathbf{P}_i \mathbf{W}(\omega_k), \mathbf{S}(\omega_k, \tau)) \} \quad (6.5)$$

bestimmt werden.

6.1.2 Entpermutation durch Korrelation mit den Eingangssignalen

Die zweite nicht-blinde Methode basiert auf der Annahme, dass die getrennten Signale mit den Originalsignalen stark korreliert sind. Um eine Entpermutation zu erreichen, müssen dazu in jedem Frequenzband alle Ausgangssignale $Y_i(\omega_k, \tau)$ mit den Originalsignalen $S_j(\omega_k, \tau)$ korreliert werden. Mit Hilfe der Korrelationsfaktoren

$$\rho_{ij} = \text{corr}(Y_i(\omega_k, \tau), S_j(\omega_k, \tau)) \quad (6.6)$$

können die Permutationen identifiziert werden. Die Funktion $\text{corr}(\cdot)$ wird im Abschnitt 6.2.1 über die Einhüllenden der komplexen Signale $Y_i(\omega_k, \tau)$ definiert.

Beispiel

Für das folgende Beispiel wurden wieder die Signale aus [76] verwendet. Bei einer FFT-Länge von 8192 wurden 4097 Frequenzbänder getrennt. Das Skalierungsproblem wurde mit Hilfe des MDP gelöst. Die Ergebnisse der beiden nicht-blinden Entpermutationsmethoden sind in Abbildung 6.1 dargestellt.

In Abbildung 6.1 (a) ist die Trennleistung der einzelnen Frequenzbänder nach der Entpermutation nach Gleichung (6.5) dargestellt. In allen Frequenzbändern ist die Trennleistung positiv, und die Gesamttrennleistung beträgt 18.44dB. In einzelnen Frequenzbändern bei tiefen Frequenzen ist die Trennleistung nahe null. Hier hat die instantane Trennung somit versagt.

In 6.1 (b) ist die Trennleistung der einzelnen Frequenzbänder nach der Korrelation mit den Eingangssignalen dargestellt. Hier haben nicht alle Frequenzbänder eine positive Trennleistung. Dies gilt besonders für die schlecht getrennten Frequenzbänder, bei denen beide Quellen in beiden Ausgängen vorhanden sind. In einzelnen Bändern führt diese Methode zu einer anderen Permutation im Vergleich zu der Entpermutation nach der Trennleistung, wie in Abbildung 6.1 (c) dargestellt. In dieser Abbildung markiert eine Null die gleiche Permutation und eine Eins eine unterschiedliche. Die Trennleistung ist mit 18.42dB im Prinzip gleich geblieben.

An diesem Beispiel sieht man, dass die Lösung des Permutationsproblems nicht eindeutig ist. So führen bereits nicht-blinde Methoden zu unterschiedlichen Permutationen, die aber nahezu gleiche Trennleistung erreichen. Das liegt daran, dass nicht getrennte Frequenzbänder bei jeder Permutation die Trennleistung

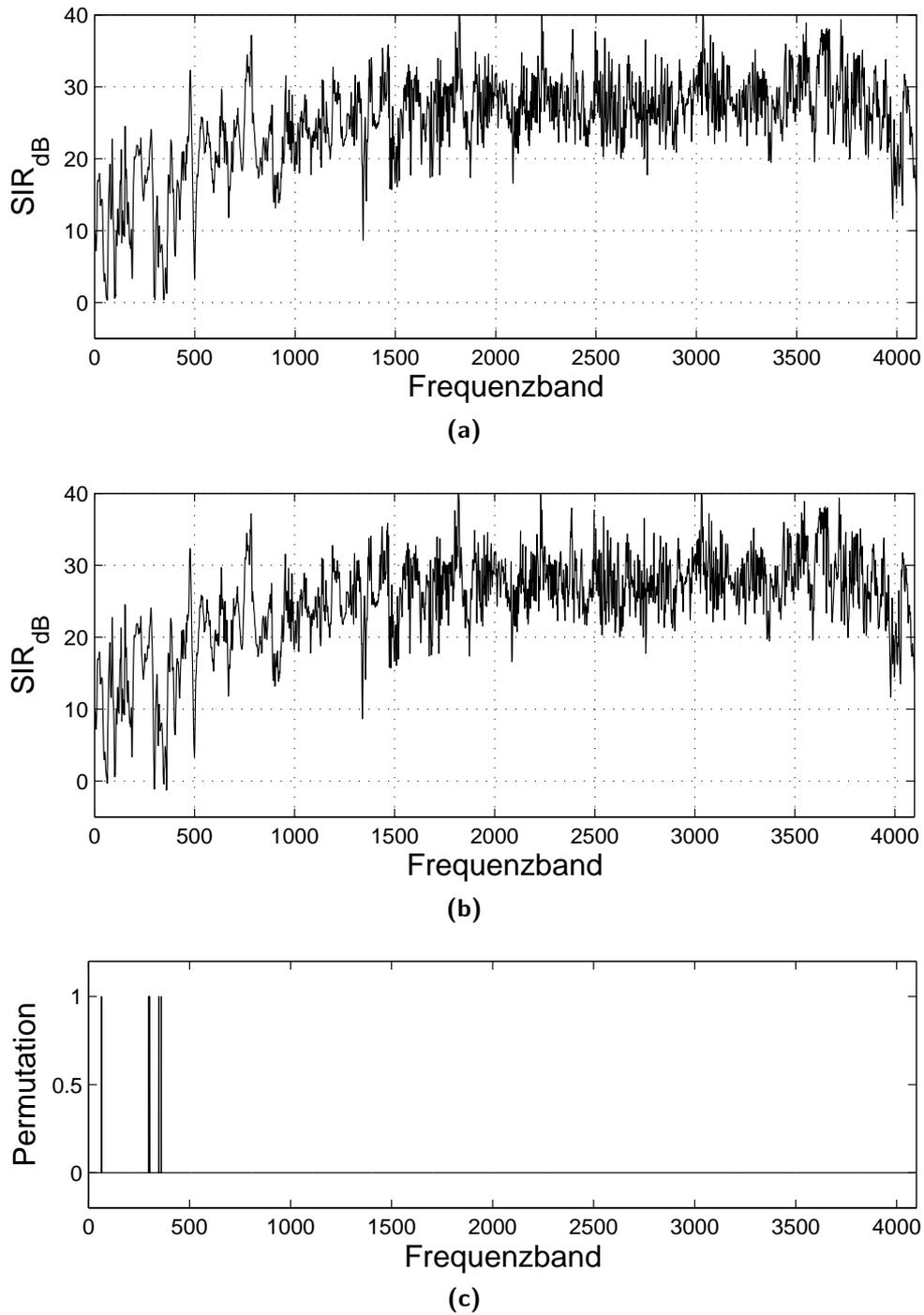


Abbildung 6.1: Beispiel für nicht-blinde Bestimmung der Permutationen. (a) Entpermutation nach Gleichung (6.5). (b) Entpermutation durch Korrelation mit den Eingangssignalen. (c) Der Unterschied zwischen den beiden Methoden. Eine Eins markiert einen Unterschied.

verkleinern. Damit ist auch eine bloße Zählung der falsch permutierten Frequenzbänder kein geeignetes Maß für die Qualität eines Entpermutationsalgorithmus.

6.2 Entpermutationsalgorithmen auf Basis getrennter Signale

Bei blinden Methoden kann die Entpermutation nur mit Hilfe der Ausgangssignale und des Entmischungssystems bestimmt werden. Im Folgenden werden Algorithmen vorgestellt, die den zeitlichen Verlauf der einzelnen Frequenzbänder benutzen.

6.2.1 Entpermutation auf Basis der Korrelation

Eine Methode, die auf der Ähnlichkeit benachbarter Frequenzbänder basiert, wurde in [32] vorgestellt. Hierzu wird angenommen, dass benachbarte Frequenzbänder miteinander stark korreliert sind. Ein Beispiel dafür ist in Abbildung 6.2 gegeben, wo der zeitliche Verlauf des Betrages von drei Frequenzbändern gezeigt ist. So haben die beiden direkt benachbarten Frequenzbänder mit den Nummern 162 und 163 einen nahezu gleichen Verlauf, was zu einer hohen Korrelation führt. Diese Annahme gilt jedoch nicht, wenn weiter entfernte Frequenzbänder verglichen werden. So ist der zeitliche Verlauf des Frequenzbandes Nummer 180 bereits deutlich anders. Dieser Verlauf kann dann unter Umständen eine größere Ähnlichkeit mit einem anderen Signal haben, so dass die Annahme der hohen Korreliertheit im Allgemeinen nur zwischen benachbarten Frequenzbändern gerechtfertigt ist.

Bei Sprachsignalen haben zusätzlich die ganzzahligen Vielfachen einer Frequenz einen ähnlichen Zeitverlauf. Mit Hilfe der *harmonischen Analyse* [73] kann dies ebenfalls zur Lösung des Permutationsproblems benutzt werden.

Berechnung der Korrelationsfaktoren

Die Korrelationsfaktoren werden auf den Einhüllenden berechnet. Mit

$$\mathbf{V}(\omega_k, \tau) = |\mathbf{Y}(\omega_k, \tau)|, \quad (6.7)$$

6.2 Entpermutationsalgorithmen auf Basis getrennter Signale

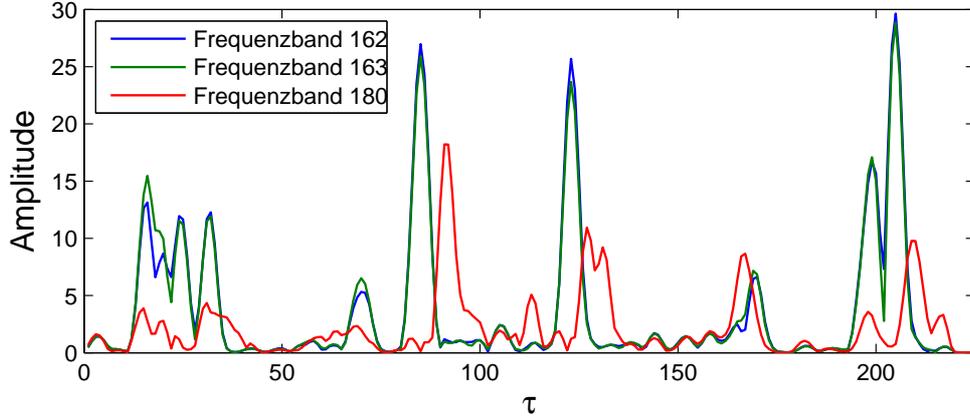


Abbildung 6.2: Der zeitliche Verlauf der Einhüllenden von drei Frequenzbändern eines Sprachsignals.

dem Betrag der STFT der getrennten Signale, ist die Korrelation zwischen zwei Frequenzbändern k und l als

$$\rho_{pq}(\omega_k, \omega_l) = \frac{\sum_{\tau=0}^{\mathcal{T}-1} V_q(\omega_k, \tau) V_p(\omega_l, \tau)}{\sqrt{\sum_{\tau=0}^{\mathcal{T}-1} V_q^2(\omega_k, \tau)} \sqrt{\sum_{\tau=0}^{\mathcal{T}-1} V_p^2(\omega_l, \tau)}} \quad (6.8)$$

definiert. Hierbei sind p und q die Indizes der getrennten Signale und \mathcal{T} die Anzahl der Frames.

Um zu entscheiden, wie zwei Quellen p und q permutiert sind, kann das Verhältnis

$$r_{pq}(\omega_k, \omega_l) = \frac{\rho_{pp}(\omega_k, \omega_l) + \rho_{qq}(\omega_k, \omega_l)}{\rho_{pq}(\omega_k, \omega_l) + \rho_{pq}(\omega_k, \omega_l)} \quad (6.9)$$

benutzt werden. Bei $r_{pq} > 1$ sind die Frequenzbänder gleich permutiert und $r_{pq} < 1$ deutet auf eine unterschiedliche Anordnung hin.

Sind mehr als zwei Quellen vorhanden, so müssen diese paarweise gegeneinander getestet werden. In diesem Fall kann es passieren, dass die Lösung nicht eindeutig ist. In [72] wurde dazu vorgeschlagen, die Faktoren $\rho_{pq}(\omega_k, \omega_l)$ zu einer Matrix

$$\mathbf{Y}(\omega_k, \omega_l) = \begin{bmatrix} \rho_{11}(\omega_k, \omega_l) & \rho_{12}(\omega_k, \omega_l) & \dots & \rho_{1N}(\omega_k, \omega_l) \\ \rho_{21}(\omega_k, \omega_l) & \rho_{22}(\omega_k, \omega_l) & \dots & \rho_{2N}(\omega_k, \omega_l) \\ \dots & \dots & \dots & \dots \\ \rho_{N1}(\omega_k, \omega_l) & \rho_{N2}(\omega_k, \omega_l) & \dots & \rho_{N,N}(\omega_k, \omega_l) \end{bmatrix} \quad (6.10)$$

zusammenzufassen und die beste Entpermutation mit

$$\mathbf{P}_{\text{KORR}}(\omega_k, \omega_l) = \underset{\mathbf{P}_i}{\operatorname{argmax}} \{ \operatorname{spur}(\mathbf{P}_i \mathbf{Y}(\omega_k, \omega_l)) \} \quad (6.11)$$

6 Das Permutationsproblem

zu bestimmen. Hierbei ist

$$\text{spur}(\mathbf{A}) = \sum_{i=1}^N a_{ii}, \quad (6.12)$$

die Summe der Diagonalelemente, also die Spur einer Matrix.

Sequentielle Entpermutation

Ein einfaches Entpermutationsschema ist die *sequentielle Entpermutation* [32, 72]. Hierbei werden zwei benachbarte Frequenzbänder anhand von $r_{pq}(\omega_k, \omega_l)$ beziehungsweise $\mathbf{P}_{\text{KORR}}(\omega_k, \omega_l)$ jeweils gleich permutiert angeordnet. Das Verfahren wird sequentiell von den unteren zu den oberen Frequenzbändern durchgeführt. Die Permutation des ersten Frequenzbandes ist dabei zufällig und bestimmt die äußere Permutation der gesamten Signale.

Beispiel

In Abbildung 6.3 ist ein Beispiel für die Koeffizienten $r_{pq}(\omega_k, \omega_l)$ bei zwei Quellen gegeben. Die Koeffizienten sind für alle Kombinationen der Frequenzbänder ω_k und ω_l berechnet worden. Rote und gelbe Bereiche bedeuten, dass die Frequenzbänder untereinander nicht permutiert sind. Die blauen und grünen Bereiche deuten auf eine Permutation hin, die aber in diesem Fall nicht vorhanden ist, da die Frequenzbänder nicht-blind entpermutiert worden sind. Noch deutlicher ist dies in Abbildung 6.4 sichtbar. Hier sind die Koeffizienten, die eine Permutation anzeigen, schwarz dargestellt. Die großen schwarz markierten Bereiche sind so zu deuten, dass die Annahme der hohen Korreliertheit über größere Bereiche nicht korrekt ist.

Da die Punkte um die Hauptdiagnale nahezu alle weiß sind, kann die sequentielle Entpermutation trotzdem zum Erfolg führen. Die Ergebnisse dieses Verfahrens sind in Abbildung 6.5 dargestellt. In diesem Beispiel sind große Bereiche korrekt entpermutiert worden. Lediglich einzelne Permutationen sind falsch bestimmt worden. Aus der sequentiellen Ausführung folgt, dass eine einzelne Permutation weitere falsche Permutationen nach sich zieht, was zu den sogenannten *Blockpermutationen* führt. Obwohl die meisten Permutationen lokal korrekt gelöst worden sind, führen Blockpermutationen zu vielen falsch permutierten Bereichen. Die Entpermutation schlägt insgesamt fehl.

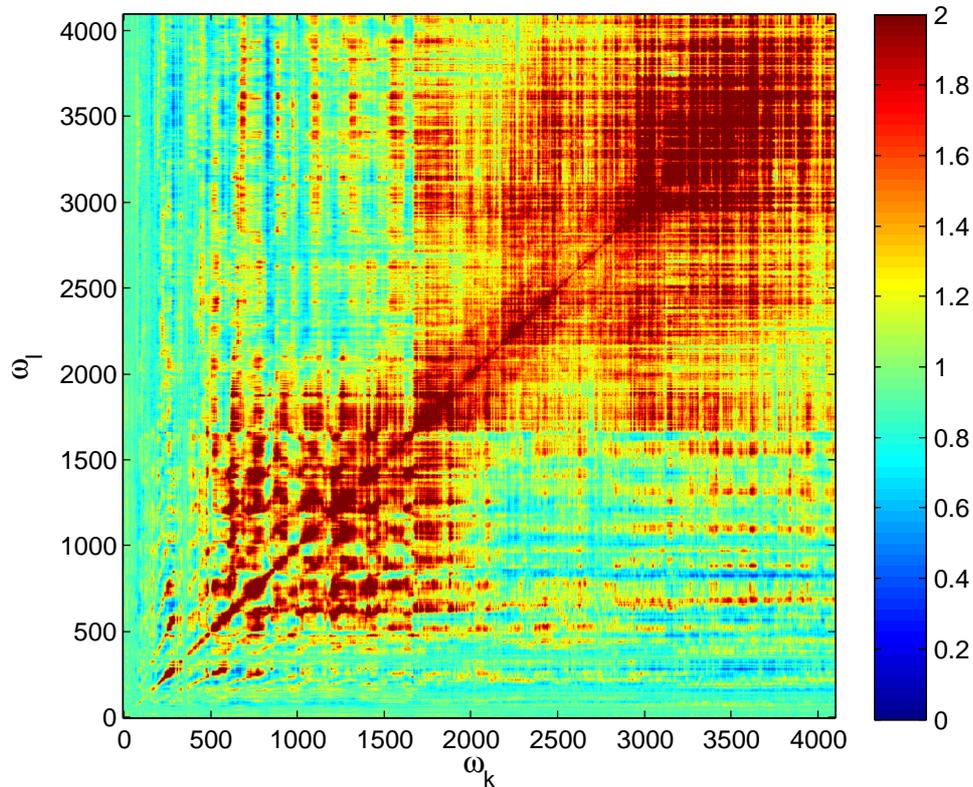


Abbildung 6.3: Beispiel für die Koeffizienten $r_{pq}(\omega_k, \omega_l)$ bei zwei entpermutierten Quellen. Die blauen und grünen Bereiche weisen auf eine Permutation hin, die aber nicht vorhanden ist.

6.2.2 Dyadisches Sortieren

Um das Problem der Blockpermutationen zu lösen, wurde in [72] ein dyadischer Sortierungsalgorithmus vorgeschlagen. Das Hauptproblem der sequentiellen Entpermutation ist, dass einzelne falsche Permutationen zu Blockpermutationen führen.

Das dyadische Verfahren versucht diesen Nachteil zu umgehen. Im ersten Schritt werden jeweils zwei benachbarte Frequenzbänder mit Hilfe von Gleichung (6.9) beziehungsweise (6.11) angeordnet. In diesem Schritt wurden die K Frequenzbänder zu $\frac{K}{2}$ Paaren zusammengefasst. Im nächsten Schritt werden diese Paare entpermutiert, die dann zu $\frac{K}{4}$ Quadrupeln führen. Diese werden dann wiederum zu Oktetten zusammengefasst. Dieses Verfahren wird so lange angewendet, bis alle Frequenzbänder zusammengefasst sind.

6 Das Permutationsproblem

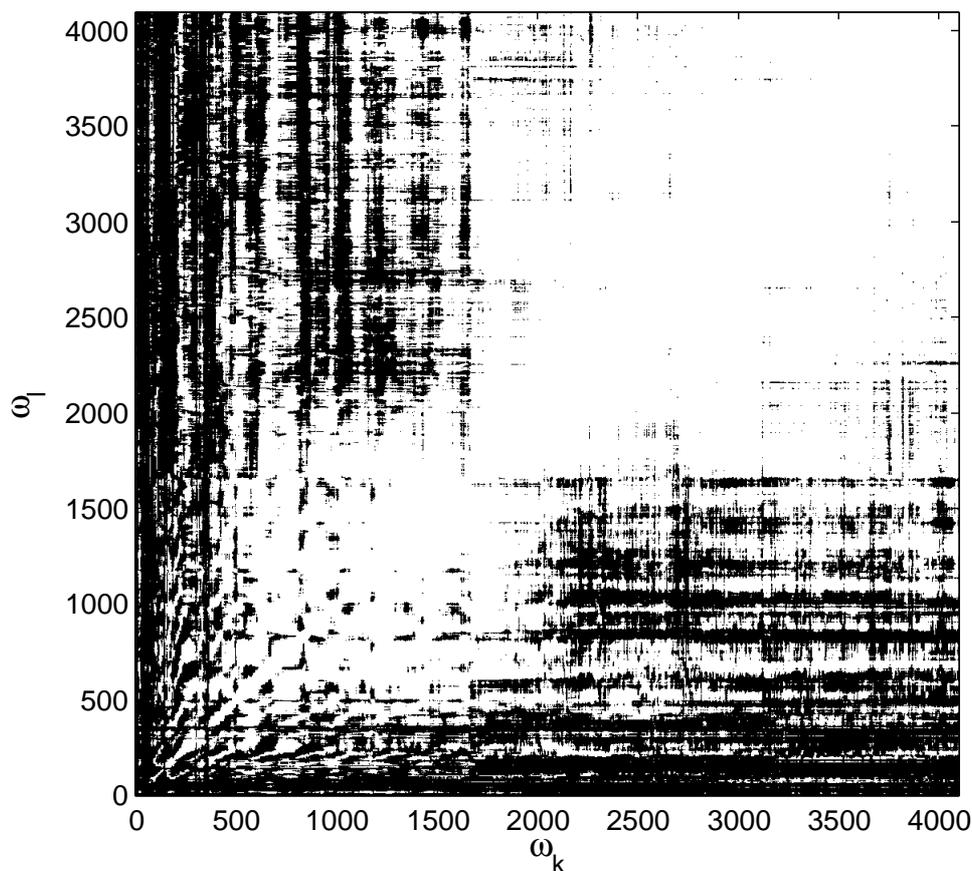


Abbildung 6.4: Koeffizienten aus Abbildung 6.3. In schwarz sind die Koeffizienten markiert, die kleiner als eins sind, und damit eine nicht vorhandene Permutation andeuten.

Der Vorteil des dyadischen Sortierens ist, dass die Gruppen mit jeweils größerer Genauigkeit entpermutiert werden können als einzelne Frequenzbänder. Dies wird erreicht, indem alle Korrelationsfaktoren zwischen allen Frequenzbändern der jeweiligen Gruppen bestimmt und diese zusammen ausgewertet werden. Für diese Auswertung kann

- der Durchschnitt aller Korrelationsfaktoren,
- das Maximum beziehungsweise das Minimum aller Korrelationsfaktoren oder
- die Anzahl der Korrelationsfaktoren größer beziehungsweise kleiner als eins

benutzt werden. Die gemeinsame Annahme bei diesen Ansätzen ist, dass einzelne zu Anfang falsch permutierte Frequenzbänder bei einer größeren Anzahl nicht

6.2 Entpermutationsalgorithmen auf Basis getrennter Signale

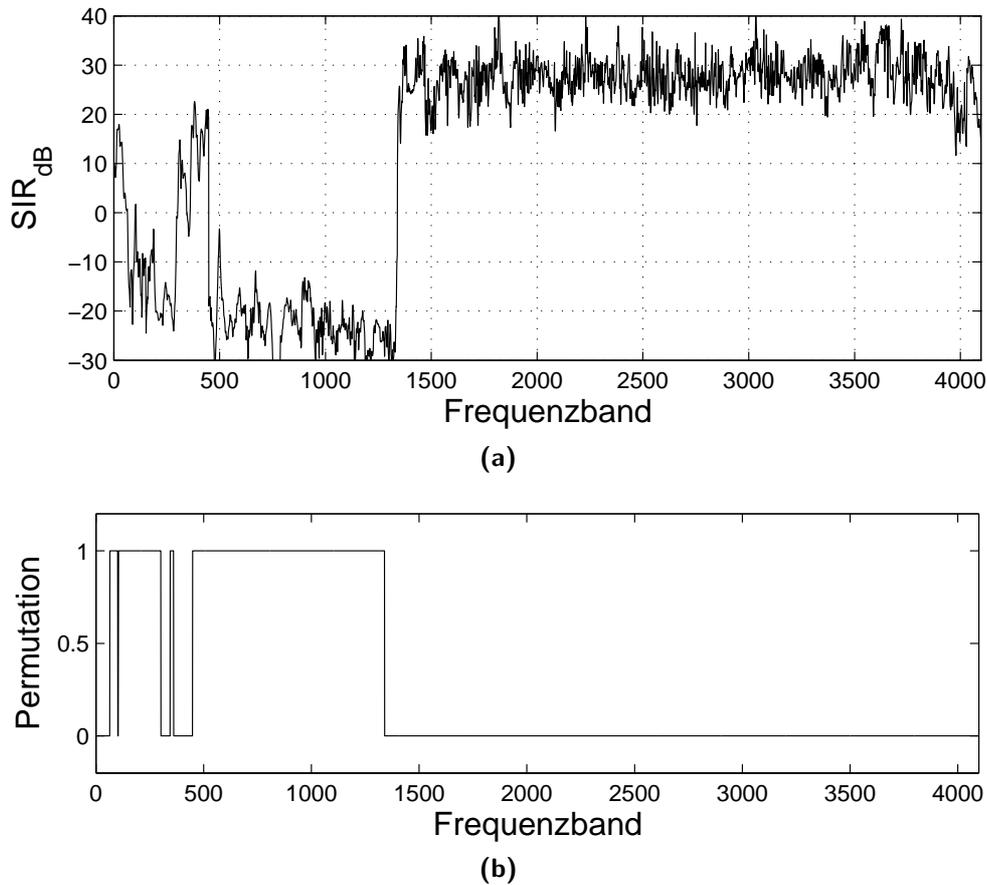


Abbildung 6.5: Die sequentielle Entpermutation. (a) Die Trennleistung in den Frequenzbändern. (b) Permutationen gegenüber einer nicht-blinden Methode.

so stark ins Gewicht fallen, so dass die Blockpermutationen trotzdem korrekt bestimmt werden können.

Beispiel

Das Ergebnis des dyadischen Sortierens ist in Abbildung 6.6 dargestellt. Im Vergleich zum sequentiellen Korrelationsverfahren konnten wesentlich größere Bereiche korrekt entpermutiert werden. So ist lediglich eine Blockpermutation im Bereich der Frequenzindizes von ungefähr 100 bis 300 falsch.

6 Das Permutationsproblem

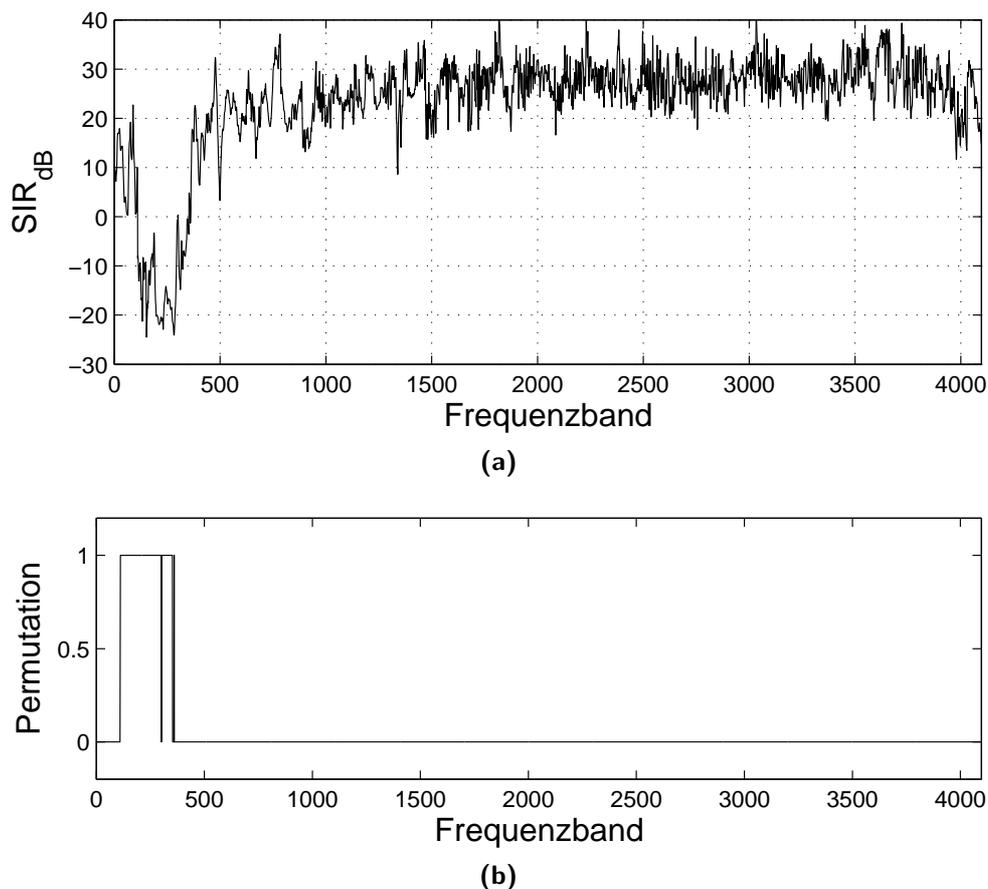


Abbildung 6.6: Dyadische Sortierung. (a) Die Trennleistung in den Frequenzbändern. (b) Permutationen gegenüber einer nicht-blinden Methode.

6.2.3 Zusammenfassung

In diesem Abschnitt wurden zwei Methoden für die Lösung des Permutationsproblems vorgestellt. Sie basieren auf der Annahme hoher Ähnlichkeit des Zeitverlaufs benachbarter Frequenzbänder. Mit Hilfe von Korrelationsfaktoren können diese benachbarten Frequenzbänder entpermutiert werden. Auch wenn der Großteil der Permutationen korrekt gelöst werden kann, so führen trotzdem einzelne falsche Zuordnungen zu Blockpermutationen. Mit Hilfe des dyadischen Sortierens können diese Probleme zwar reduziert, aber nicht vollständig gelöst werden.

6.3 $\alpha\beta$ -Algorithmus

In diesem Abschnitt wird eine neue Methode für die Lösung des Permutationsproblems vorgestellt, die bereits in [48, 50, 52] in Teilen veröffentlicht wurde. Der zentrale Ansatz ist es, die einzelnen Frequenzbänder mit Hilfe der GGD nach Gleichung (3.80) zu modellieren und für die Entpermutation die Kontinuität der Parameter α und β über die Frequenzbänder zu benutzen.

In [79] wurde gezeigt, dass der β -Parameter bei der Mittelung über eine genügend große Anzahl von Sprachsignalen nahezu konstant für alle Frequenzbänder ist. Gleichzeitig wurde aber auch gezeigt, dass er für kurze Signale sehr unterschiedliche Werte annehmen kann. In Abbildung 6.7 (a) ist ein Verlauf $\beta(\omega_k)$ dargestellt. Das Besondere dieses Parameters ist, dass der Wert von β sich in den benachbarten Frequenzbändern kaum ändert, aber gleichzeitig für unterschiedliche Signale in den meisten Frequenzbändern deutlich unterschiedlich ist, wie in Abbildung 6.7 (b) an einem Ausschnitt genauer dargestellt ist. Dies kann nun genutzt werden, um Bereiche (Cluster) zu bestimmen, in denen einzelne Frequenzbänder so permutiert werden können, dass die Werte für β den einzelnen Signalen eindeutig zugeordnet werden können. In diesen Bereichen wird durch diese Zuordnung das Permutationsproblem gelöst.

Mit der beschriebenen Methode wird das Permutationsproblem zwar nicht in allen Frequenzbändern gelöst, aber es ergibt sich eine erhebliche Vereinfachung. So müssen nun nicht mehr mehrere tausend Frequenzbänder, sondern nur noch die eindeutig bestimmten Cluster entpermutiert werden. In typischen Fällen ist die Anzahl der Cluster um ein bis zwei Größenordnungen kleiner als die der Frequenzbänder. Für die Entpermutation der Cluster kann die Korrelationsmethode des dyadischen Sortierens benutzt werden. Das dyadische Sortieren liefert dann gute Ergebnisse, wenn die zu entpermutierenden Bereiche selbst keine falschen Permutationen enthalten, was durch das Clustern sichergestellt wird. Nach der Entpermutation der Cluster können dann schließlich die restlichen Frequenzbänder, die nicht geclustert wurden, mit der Korrelationsmethode entpermutiert werden.

Dieses Verfahren kann analog für den Parameter α angewandt werden. Die Frequenzbänder werden dann unabhängig von β nach einem ähnlichen Muster geclustert. In Abbildung 6.8 (a) ist der Verlauf des Parameters $\alpha(\omega_k)$ für die einzelnen Frequenzbänder dargestellt. Da $\alpha(\omega_k)$ und $\beta(\omega_k)$ unterschiedliche Verläufe haben, wie im Detail in den Abbildungen 6.7 (a) und 6.8 (b) dargestellt, sind die jeweiligen Bereiche, die geclustert werden können, unterschiedlich. Durch Nutzung überlappender Bereiche können diese Cluster dann in größere

6 Das Permutationsproblem

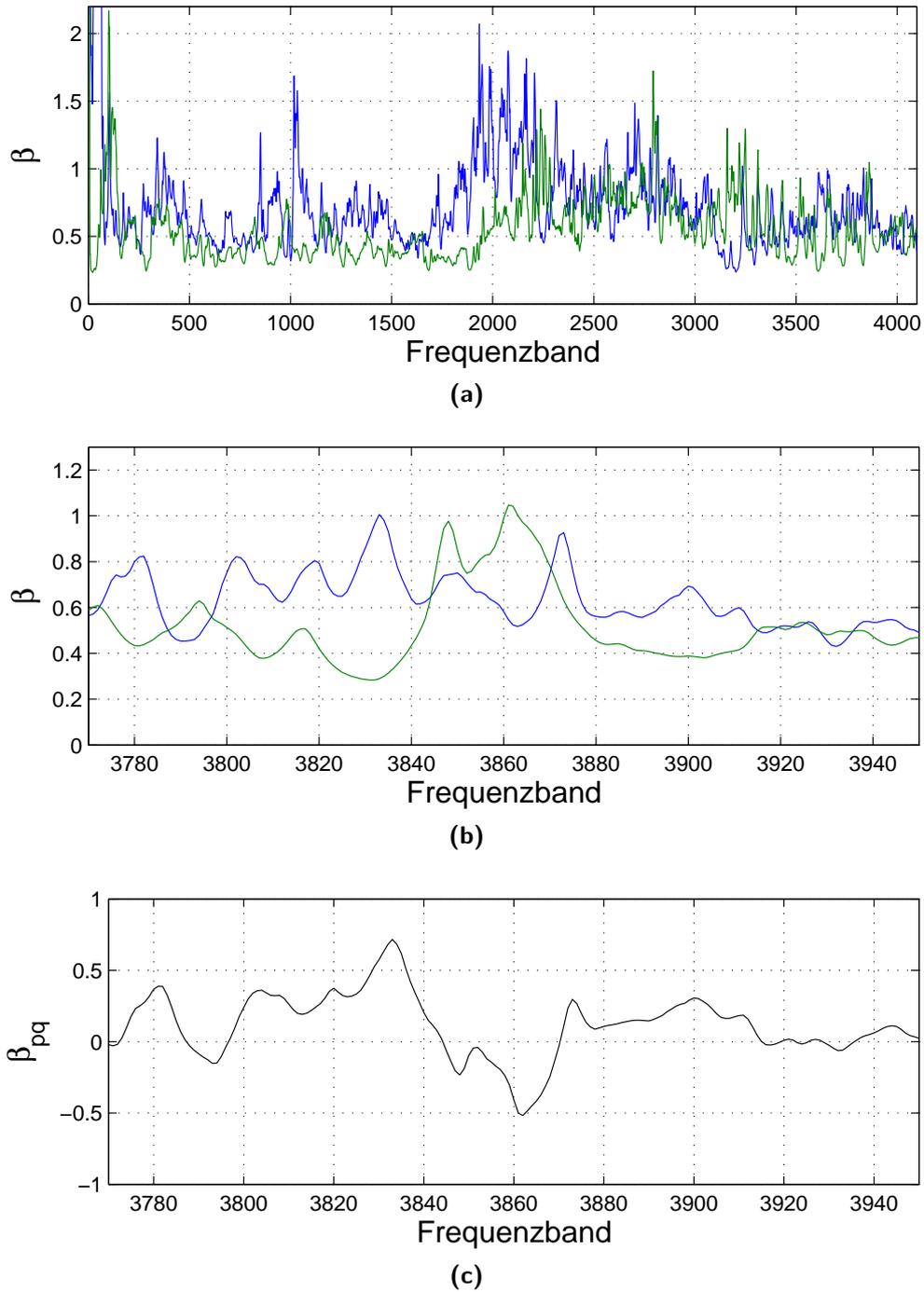


Abbildung 6.7: Beispiel für den Verlauf des β -Parameters der GGD in den einzelnen Frequenzbändern für zwei Signale. (a) Der Verlauf über alle Frequenzbänder. (b) Ein Ausschnitt. (c) Die Differenz der Werte.

zusammengefasst werden, wodurch die Robustheit des Algorithmus erheblich verbessert werden kann [50].

Eine bessere Variante, die in diesem Abschnitt vorgestellt wird, ist es, ein gemeinsames Kriterium für das Clustern mit Hilfe beider Parameter aufzustellen. Im Folgenden wird gezeigt, wie die Cluster für beide Parameter gebildet werden können, sowie wie ein gemeinsames Kriterium hergeleitet werden kann.

6.3.1 Clusterberechnung

Der erste Schritt des Entpermutationsalgorithmus ist die Berechnung der Cluster. Anschließend wird eine neue Methode für beide Parameter vorgestellt, die es erlaubt, beide Parameter in einem gemeinsamen Kriterium zusammenzufassen.

Bestimmung und Vorverarbeitung der GGD-Parameter

Der erste Schritt ist die Bestimmung der GGD-Parameter für alle Frequenzbänder und für alle Signale. Dies kann mit Hilfe des Schätzers aus Kapitel 3.3.4 durchgeführt werden. Da nicht die absoluten Verläufe, sondern die Abstände zwischen den Werten interessant sind, werden die Differenzen

$$\beta_{pq}(\omega_k) = \beta_p(\omega_k) - \beta_q(\omega_k) \quad (6.13)$$

betrachtet. Ein möglicher Verlauf von $\beta_{pq}(\omega_k)$ wurde bereits in Abbildung 6.7 (c) gezeigt. Der Parameter α wird aufgrund eines anderen Wertebereiches wie folgt in logarithmischer Form verwendet:

$$\alpha_{pq}(\omega_k) = \log(\alpha_p(\omega_k)) - \log(\alpha_q(\omega_k)). \quad (6.14)$$

Der Beispielsverlauf für $\alpha_{pq}(\omega_k)$ für den gleichen Bereich wie in Abbildung 6.7 (c) ist in Abbildung 6.8 (c) dargestellt.

Entpermutationskriterium

Ausgehend von der Ähnlichkeit der Parameter $\beta(\omega_k)$ und $\alpha(\omega_k)$ in benachbarten Frequenzbändern gilt auch für die Differenzen $\beta_{pq}(\omega_k)$ und $\alpha_{pq}(\omega_k)$, dass die benachbarten Werte sich nur wenig unterscheiden. Diese Eigenschaft kann benutzt werden, um die Permutationen in benachbarten Frequenzbändern zu bestimmen.

6 Das Permutationsproblem

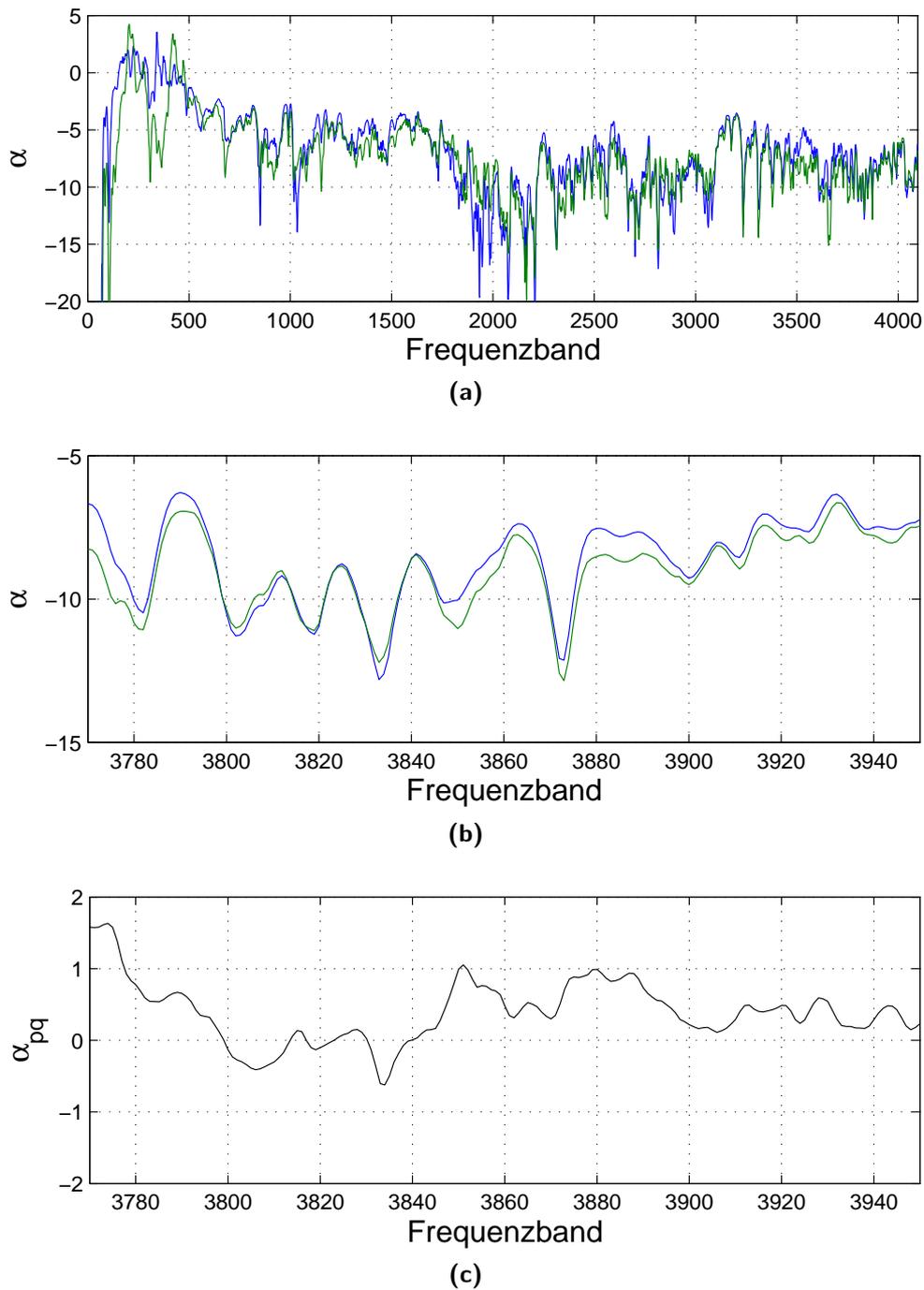


Abbildung 6.8: Beispiel für den Verlauf des α -Parameters der GGD in den einzelnen Frequenzbändern für zwei Signale. (a) Der Verlauf über alle Frequenzbänder. (b) Ein Ausschnitt. (c) Die Differenz der Werte.

Eine Permutation zwischen zwei Frequenzbändern ω_k und ω_{k+1} äußert sich darin, dass der Verlauf von $\beta_{pq}(\omega)$ an dieser Stelle nicht glatt ist und einen Sprung mit einem Vorzeichenwechsel hat. Es ist sofort ersichtlich, dass die Entpermutation nur an den Stellen eindeutig bestimmt werden kann, an denen sich $\beta_{pq}(\omega)$ deutlich von null unterscheidet.

Für die Bestimmung der angesprochenen Glattheit kann ein Prädiktor verwendet werden. So kann aus dem Cluster der bereits entpermutierten Frequenzbänder $\beta(\omega_l)$ bis $\beta(\omega_k)$, $l < k$, eine Vorhersage des Parameters $\beta_{\text{pred}_{pq}}(\omega_{k+1})$ des nächsten Frequenzbandes erstellt werden.

Liegt keine Permutation zwischen dem Cluster der bereits entpermutierten und dem folgenden Frequenzband vor, so ist der Fehler

$$\beta_{d_1}(\omega_{k+1}) = \beta_{\text{pred}_{pq}}(\omega_{k+1}) - \beta_{pq}(\omega_{k+1}) \quad (6.15)$$

betragsmäßig kleiner als der Fehler

$$\beta_{d_2}(\omega_{k+1}) = \beta_{\text{pred}_{pq}}(\omega_{k+1}) - \beta_{qp}(\omega_{k+1}), \quad (6.16)$$

der sich bei einer Vertauschung ergibt. Ist $\beta_{d_1}(\omega_{k+1})$ betragsmäßig größer als $\beta_{d_2}(\omega_{k+1})$, so liegt eine Permutation vor. Diese beiden Vergleiche können mit

$$r_k = \frac{|\beta_{d_1}(\omega_{k+1})|}{|\beta_{d_2}(\omega_{k+1})|} \quad (6.17)$$

zusammengefasst werden. Bei

$$r_k < 1 \quad (6.18)$$

ist das Frequenzband ω_{k+1} korrekt ausgerichtet, und

$$r_k > 1 \quad (6.19)$$

deutet auf eine Permutation hin. Da die Schätzung des Prädiktors fehlerbehaftet ist, sollte der Vergleich nicht absolut, sondern in einem Toleranzbereich durchgeführt werden. Wird r_k mit

$$r_k < \frac{1}{\rho} \quad (6.20)$$

und

$$r_k > \rho, \quad (6.21)$$

mit $\rho > 1$ verglichen, so gibt es ein Intervall $[\frac{1}{\rho}, \rho]$, in dem keine Aussage über die Permutation gemacht wird. Das bedeutet, das Frequenzband ω_{k+1} kann nicht

6 Das Permutationsproblem

zum Cluster hinzugefügt werden und seine Zuordnung kann erst in der nächsten Stufe mit Hilfe der Korrelationsmethode durchgeführt werden. Liegt r_k außerhalb dieses Intervalls, so kann die Permutation mit großer Sicherheit korrekt bestimmt und dieses Frequenzband mit der richtigen Permutation zum Cluster hinzugefügt werden. Danach wird das Verfahren für das nächste Frequenzband durchgeführt.

Lineare Prädiktoren

Die Bestimmung des prädizierten Wertes $\beta_{\text{pred}_{pq}}(\omega_{k+1})$ kann mit Hilfe eines linearen Prädiktors $a(n)$ der Länge L_a durchgeführt werden. Je länger $a(n)$ ist, desto besser können die Werte vorhergesagt werden, das heißt, der mittlere Fehler $E\{|\beta_{d_1}(\omega_{k+1})|^2\}$ wird kleiner. In Tabelle 6.1 ist diese Abhängigkeit für $n = 2, \dots, 7$ dargestellt. Es ist ersichtlich, dass bereits kurze Prädiktoren sehr gut in der Lage sind, die Werte vorherzusagen. Diese Prädiktoren haben eine gemeinsame Struktur, bei der die ersten beiden Koeffizienten dominant sind. Da kurze Prädiktoren besser generalisieren, das heißt auch für andere Signale gute Vorhersagen liefern, wird im Folgenden der Prädiktor der Länge zwei benutzt. Mit den Koeffizienten

$$a(1) = 2 \quad \text{und} \quad a(2) = -1 \quad (6.22)$$

führt dieser Prädiktor zu einer linearen Extrapolation, mit dem die Fehler $\beta_{d_1}(\omega_{k+1})$ und $\beta_{d_2}(\omega_{k+1})$ als

$$\beta_{d_1}(\omega_{k+1}) = -\beta_{pq}(\omega_{k-1}) + 2\beta_{pq}(\omega_k) - \beta_{pq}(\omega_{k+1}) \quad (6.23)$$

und

$$\beta_{d_2}(\omega_{k+1}) = -\beta_{qp}(\omega_{k-1}) + 2\beta_{qp}(\omega_k) - \beta_{qp}(\omega_{k+1}) \quad (6.24)$$

berechnet werden können. In der Abbildung 6.9 ist der Verlauf von $\beta_{pq}(\omega_k)$ in blau und der Fehler $\beta_{d_1}(\omega_{k+1})$ in grün dargestellt. Die Vorhersage mit dem Prädiktor der Länge zwei ist gut, da der Fehler deutlich kleiner als $\beta_{pq}(\omega_k)$ ist.

Die gleiche Herleitung kann für den Parameter $\alpha(\omega_k)$ gemacht werden. So wird ein analoges Entscheidungskriterium

$$r_k = \frac{|\alpha_{d_1}(\omega_{k+1})|}{|\alpha_{d_2}(\omega_{k+1})|} \quad (6.25)$$

mit

$$\alpha_{d_1}(\omega_{k+1}) = -\alpha_{pq}(\omega_{k-1}) + 2\alpha_{pq}(\omega_k) - \alpha_{pq}(\omega_{k+1}) \quad (6.26)$$

und

$$\alpha_{d_2}(\omega_{k+1}) = -\alpha_{pq}(\omega_{k-1}) + 2\alpha_{pq}(\omega_k) - \alpha_{qp}(\omega_{k+1}) \quad (6.27)$$

benutzt. Der Verlauf von $\alpha_{pq}(\omega_k)$ und des Prädiktionsfehlers $\alpha_{d_1}(\omega_k)$ ist in Abbildung 6.10 dargestellt.

Gemeinsames Kriterium

Mit den Kriterien aus den Gleichungen (6.17) und (6.25) können Cluster für beide Parameter bestimmt werden. Da die Verläufe von $\alpha_{pq}(\omega_k)$ und $\beta_{pq}(\omega_k)$ unterschiedlich sind, sind die resultierenden Cluster unterschiedlich. Durch überlappende Bereiche können die jeweiligen Cluster zusammengefasst werden und so größere entpermutierte Bereiche bilden. Es ist jedoch auch möglich, ein gemeinsames Kriterium zu erstellen. Dazu müssen die beiden Parameter zu einem Vektor

$$\mathbf{z}_{pq}(\omega) = \left[\frac{\alpha_{pq}(\omega)}{\sigma_\alpha}, \frac{\beta_{pq}(\omega)}{\sigma_\beta} \right]^T \quad (6.28)$$

zusammengefasst werden. Hierbei sind σ_α und σ_β jeweils die Standardabweichung von $\alpha_{pq}(\omega)$ beziehungsweise $\beta_{pq}(\omega)$. Diese Normierung ist notwendig, damit beide Parameter gleichen Einfluss auf das Entpermutationskriterium haben. Analog zu den beiden bereits beschriebenen Verfahren wird ein linearer Prädiktor $\mathbf{z}_{\text{pred}_{pq}}(\omega)$ ebenfalls mit der Länge zwei entworfen und die beiden Fehler für den korrekten und den permutierten Fall mit

$$\Delta_1(\omega_{k+1}) = \|\mathbf{z}_{pq}(\omega_{k-1}) + 2\mathbf{z}_{pq}(\omega_k) - \mathbf{z}_{pq}(\omega_{k+1})\|_{\ell_2} \quad (6.29)$$

und

$$\Delta_2(\omega_{k+1}) = \|\mathbf{z}_{qp}(\omega_{k-1}) + 2\mathbf{z}_{qp}(\omega_k) - \mathbf{z}_{qp}(\omega_{k+1})\|_{\ell_2} \quad (6.30)$$

berechnet. Die Entscheidung über die Permutation wird analog mit

$$r_k = \frac{\Delta_1(\omega_{k+1})}{\Delta_2(\omega_{k+1})} \quad (6.31)$$

und dem Toleranzbereich, welcher wie in den Gleichungen (6.20) und (6.21) bestimmt ist, getroffen.

Cluster Entpermutation

Nachdem die Cluster bestimmt worden sind, müssen diese gegeneinander noch entpermutiert werden. Dies kann mit Hilfe der Korrelationsmethode durchgeführt werden, wobei die Methoden des dyadischen Sortierens sich am besten

Tabelle 6.1: Der Zusammenhang zwischen der Länge des linearen Prädiktors L_a und dem Fehler $E\{|\beta_{d_1}(\omega_{k+1})|^2\}$.

L_a	2	3	4	5	6	7
Fehler	0.00408	0.00354	0.00348	0.00347	0.00347	0.00347

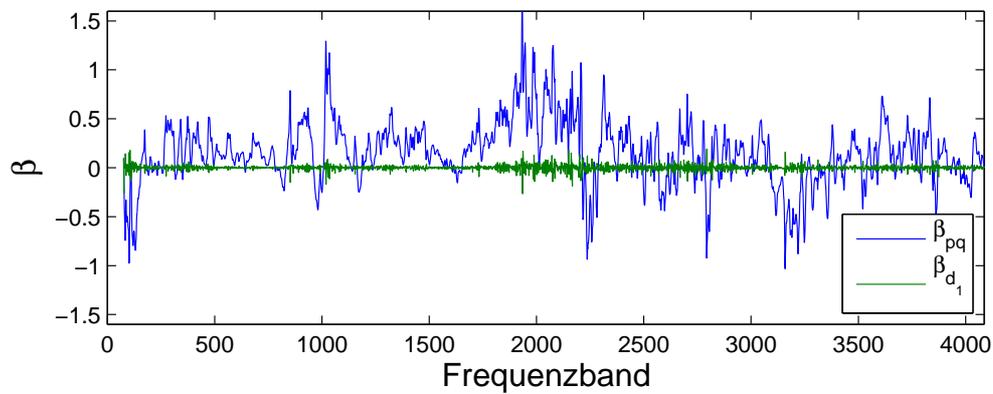


Abbildung 6.9: Der Verlauf von $\beta_{pq}(\omega_k)$ in blau nach Gleichung (6.13) und in grün der Prädiktionsfehler $\beta_{d_1}(\omega_k)$ nach Gleichung (6.23).

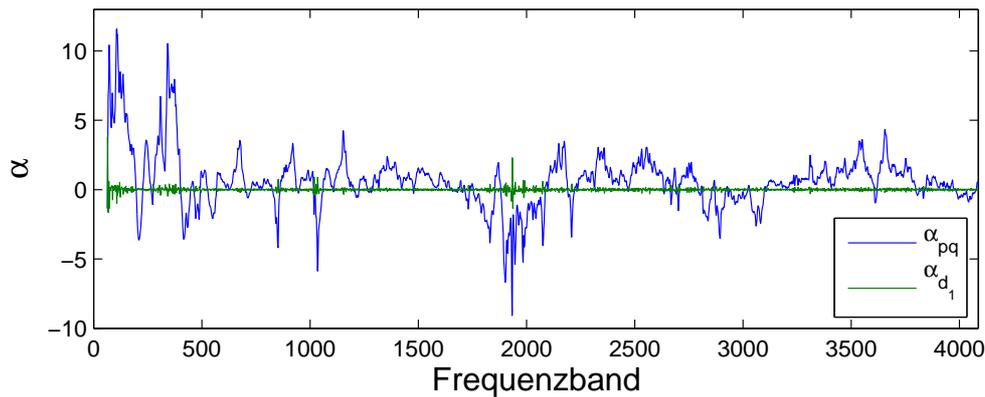


Abbildung 6.10: Der Verlauf von $\alpha_{pq}(\omega_k)$ in blau nach Gleichung (6.14) und in grün der Prädiktionsfehler $\alpha_{d_1}(\omega_k)$ nach Gleichung (6.26).

Tabelle 6.2: Vergleich der Clustergrößen der $\alpha\beta$ -Verfahrens. Nur Cluster mit mehr als acht Frequenzbändern wurden gezählt.

	Anzahl	Geclusterte Frequenzbänder	Durchschnittliche Clustergröße
α -Cluster	106	3371	31.80
β -Cluster	98	3159	32.23
Gemeinsames Verfahren	71	3706	52.20

eigenen. Als letzten Schritt müssen noch die Frequenzbänder entpermutiert werden, die zu keinem Cluster zugeordnet worden sind. Auch hier kann die Korrelationsmethode benutzt werden, wobei die einzelnen Frequenzbänder gegen zwei Cluster getestet werden können. Durch die Korrelation mit dem jeweils höher und niedriger liegenden Cluster ist eine sicherere Aussage über die Permutation möglich, als durch eine einzelne Korrelation mit dem benachbarten Frequenzband.

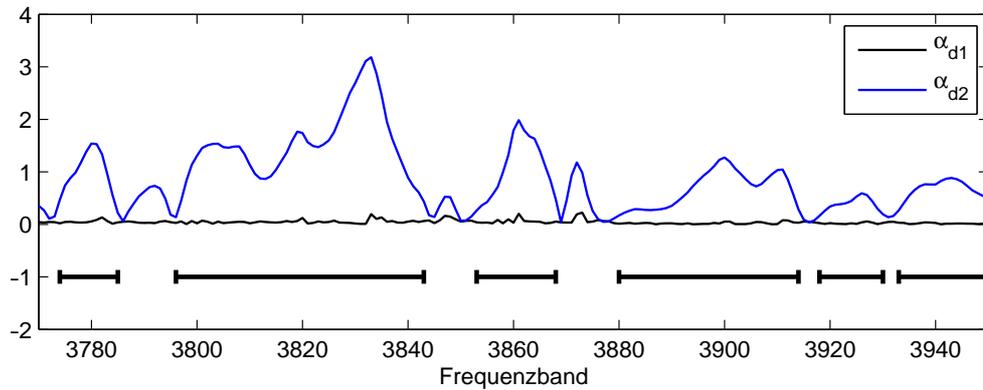
Beispiel

Die Signale des letzten Beispiels wurden ebenfalls mit Hilfe des $\alpha\beta$ -Verfahrens entpermutiert. In Abbildung 6.11 sind die Ergebnisse der Clusterbildung der drei vorgestellten Verfahren dargestellt. In Abbildung 6.11 (a) sind jeweils die beiden Fehlersignale $\alpha_{d_1}(\omega_{k+1})$ und $\alpha_{d_2}(\omega_{k+1})$ dargestellt, sowie die Bereiche, in denen die Permutation durch das Cluster-Verfahren korrekt bestimmt werden konnte. Analog dazu sind in Abbildung 6.11 (a) die Cluster dargestellt, die mit Hilfe des β -Parameters bestimmt wurden. Schließlich ist in 6.11 (c) das gemeinsame Verfahren dargestellt. Im Vergleich zu den beiden einfachen Verfahren konnten mehr Frequenzbänder zusammengefasst werden, und die Cluster werden größer.

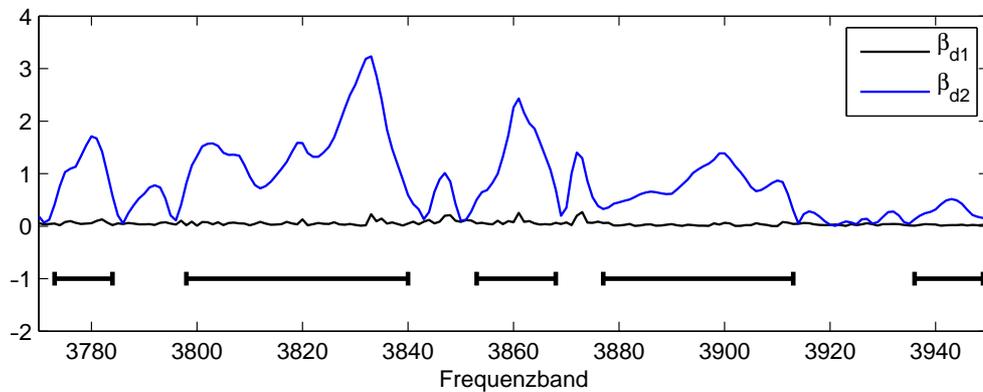
Diese Beobachtung spiegelt sich in der Anzahl und der Größe der Cluster aus Tabelle 6.2 wider. So konnten durch das gemeinsame Verfahren in 71 Clustern insgesamt 3706 der 4097 Frequenzbänder zusammengefasst werden.

Alle Cluster konnten korrekt gegeneinander entpermutiert werden, und alle, bis auf einige wenige Frequenzbänder in den ganz tiefen Frequenzen, konnten

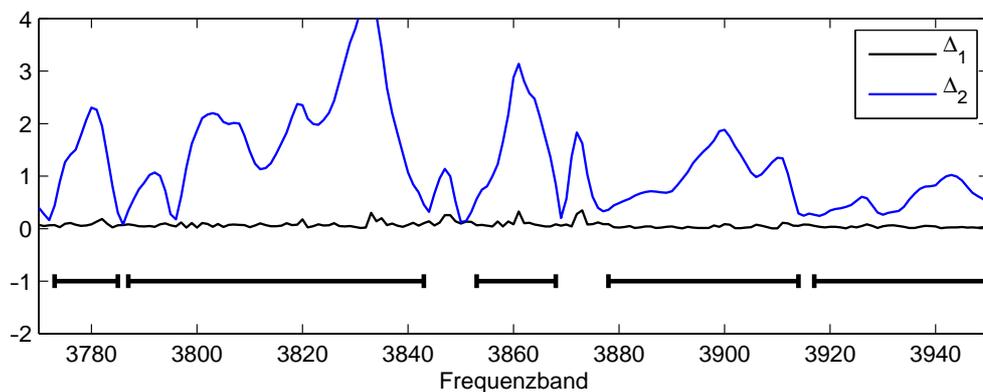
6 Das Permutationsproblem



(a)



(b)



(c)

Abbildung 6.11: Die Verläufe der Fehlersignale sowie die geclusterten Bereiche. (a) Berechnung für $\alpha_{d1}(\omega_{k+1})$ und $\alpha_{d2}(\omega_{k+1})$. (b) Berechnung für $\beta_{d1}(\omega_{k+1})$ und $\beta_{d2}(\omega_{k+1})$. (c) Berechnung für das gemeinsame Kriterium mit $\Delta_1(\omega_{k+1})$ und $\Delta_2(\omega_{k+1})$.

Tabelle 6.3: Vergleich der Trennleistung der verschiedenen Algorithmen

	2 Kanäle	3 Kanäle	4 Kanäle
$\alpha\beta$ Algorithmus	18.4	12.3	8.3
$\alpha\beta$ Algorithmus mit Filterformung	26.8	18.1	13.2
Sequentielles Entpermutieren	3.1	1.4	0
Dyadisches Sortieren	14.3	1.9	0
DOA - Algorithmus	17.3	12.4	9.2
Nicht-blind	18.4	13.6	9.9

ebenfalls korrekt zugeordnet werden, wie in Abbildung 6.12 dargestellt. Das Frequenzband Nummer 364 ist eines der Bänder, welches bereits die nicht-blinden Methoden unterschiedlich zuordnen. Die Frequenzbänder von null bis 100 enthalten keine Anteile der Sprachsignale, so dass die Permutation hier beliebig ist, ohne einen nennenswerten Einfluss auf die Gesamttrennleistung zu haben.

In Tabelle 6.3 sind die Ergebnisse der in diesem Kapitel vorgestellten Algorithmen zusammengefasst. Zusätzlich ist noch der sogenannte DOA-Algorithmus [73] aufgeführt, der die Eigenschaften der Entmischungsmatrizen zur Entpermutation benutzt. Die Voraussetzung für diesen Algorithmus ist allerdings die Kenntnis der Positionen der Mikrofone, so dass es sich nicht um einen vollständig blinden Algorithmus handelt. Für diesen Vergleich wurde das Skalierungsproblem mit Hilfe des MDP gelöst. Zusätzlich ist für den $\alpha\beta$ -Algorithmus noch die Lösung mit der Filterformung gegeben. In allen Fällen kann durch das Filterformen die Trennleistung zusätzlich verbessert werden. Wie man sieht, ist der $\alpha\beta$ -Algorithmus bei zwei Kanälen genauso gut wie der nicht-blinde Algorithmus. Bei mehreren Kanälen konnte kein vorgestellter Algorithmus alle Frequenzbänder entpermutieren. Nur der $\alpha\beta$ - und der DOA-Algorithmus waren in der Lage, eine Trennung zu erzielen. Beide Algorithmen liefern ähnlich gute Ergebnisse, obwohl sie auf ganz unterschiedlichen Ansätzen basieren.

6 Das Permutationsproblem

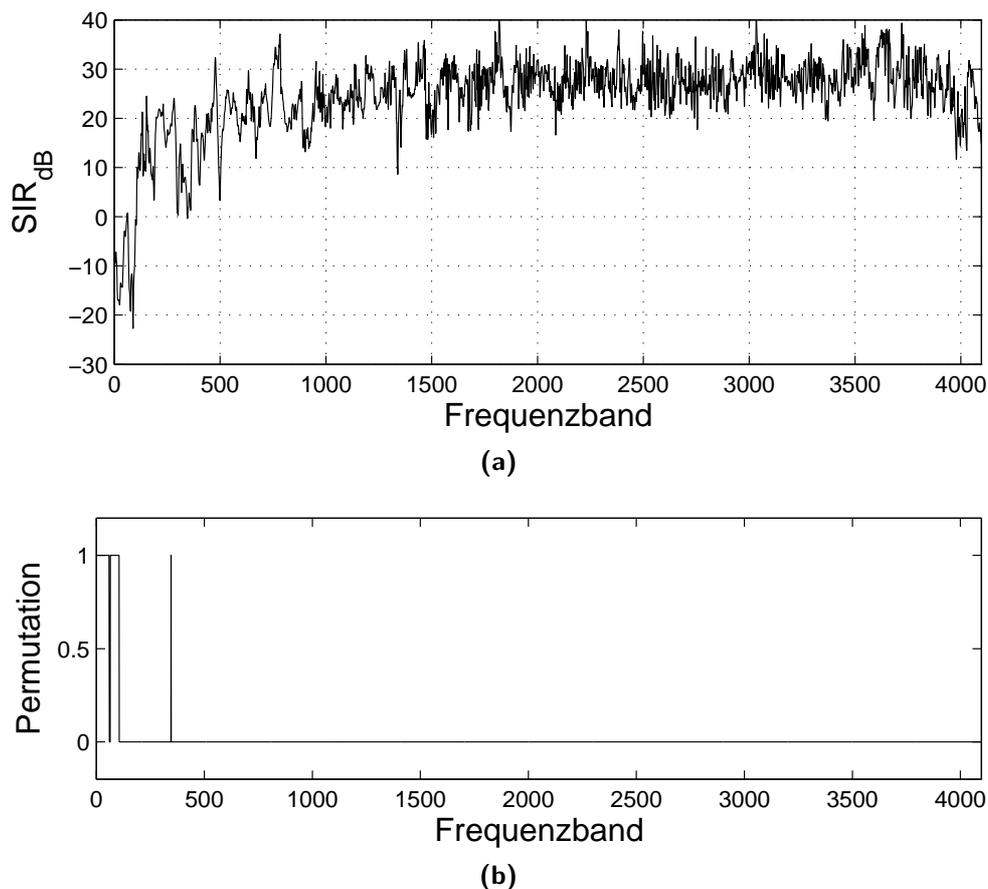


Abbildung 6.12: Die Entpermutation mit der $\alpha\beta$ -Methode. (a) Die Trennleistung in den Frequenzbändern. (b) Permutationen gegenüber einer nicht-blinden Methode.

6.4 Zusammenfassung

In diesem Kapitel wurde ein neuer Algorithmus für die Lösung des Permutationsproblems vorgestellt. Er basiert auf der Modellierung der einzelnen Frequenzbänder mit Hilfe der GGD und nutzt die kleine Änderung der Parameter in den benachbarten Frequenzbändern. An einem Beispiel wurde demonstriert, dass der neue Algorithmus eine sehr gute Leistung zeigt. Keine der Methoden, die auf der Statistik der getrennten Signale basiert, hat die Fähigkeit, so gut zu entpermutieren. Lediglich der Ansatz, der die Entmischungsmatrizen auswertet und nicht vollständig blind ist, ist in der Lage, ähnlich gute Ergebnisse zu liefern.

Kapitel 7

Zusammenfassung und Ausblick

In dieser Arbeit wurde gezeigt, wie unbekannte Mischungen von Signalen wieder getrennt werden können. Die Motivation dazu ist das Cocktail-Party-Problem, welches nicht nur rein vom theoretischen Standpunkt her interessant ist, sondern auch viele praktische Anwendungen hat.

Die besondere Schwierigkeit des Cocktail-Party-Problems ist das konvolutive Mischungssystem. Um solcherart gemischte Signale wieder zu trennen, muss wiederum ein konvolutives Entmischungssystem entworfen werden. Da dies im Zeitbereich sehr aufwändig ist, wird das Problem im Zeit-Frequenz-Bereich gelöst. So wird das konvolutive Problem in viele instantane, voneinander unabhängige Probleme transformiert, welche mit der Annahme der statistischen Unabhängigkeit der Signale mit Hilfe der Independent Component Analysis gelöst werden können.

Dieser Vorteil wird jedoch mit dem Nachteil der beliebigen Skalierung und Permutation in jedem Frequenzband erkauft. Diese beiden Probleme müssen gelöst werden, da das Verfahren sonst lediglich gefilterte oder sogar nicht getrennte Signale liefert.

In dieser Arbeit wurden neue Ansätze zur Lösung dieses Skalierungs- und Permutationsproblems vorgestellt. Eine neue Lösung des Skalierungsproblems zielt darauf ab, die Trennleistung zu maximieren. Dies wird mit Hilfe einer blinden Schätzung der Trennleistung in den einzelnen Frequenzbändern erreicht. Zwei weitere neue Ansätze formen die Entmischungsfiler so, dass diese einem Dirac-Impuls oder typischen Raumimpulsantworten ähnlich werden. Dieser Ansatz ist psychoakustisch motiviert und hat das Ziel, die Signale möglichst angenehm für das menschliche Ohr klingen zu lassen. An dieser Stelle stellen sich weitere Fragen, die in dieser Arbeit nicht behandelt worden sind, wie zum Beispiel, wie

7 Zusammenfassung und Ausblick

psychoakustische Modelle zur Modellierung und Messung verwendet werden können.

Das zweite zentrale Problem der Lösung im Zeit-Frequenz-Bereich ist das Permutationsproblem. Hierzu wurde eine neue Methode basierend auf den Eigenschaften der entmischten Signale vorgestellt. Sie basiert auf einer Modellierung mit Hilfe der generalisierten Gaussdichte und nutzt die kleinen Unterschiede der Parameter in benachbarten Frequenzbändern. Bei diesem Ansatz werden die Eigenschaften des Entmischungssystems nicht beachtet. Es existieren weitere gute Algorithmen, die genau dies für das Entpermutationskriterium benutzen. Eine Kombination der beiden Ansätze verspricht ein sehr robustes Verfahren.

Abbildungsverzeichnis

1.1	Das Cocktail Party Problem.	3
2.1	Beispiel für eine Mischung von Signalen. (a) Drei Sprachsignale und (b) eine lineare Mischung.	10
2.2	Beispiel für ein allgemeines Misch- und Entmischungssystem. Auf der linken Seite sind die unbekanntenen Größen dargestellt. Auf der rechten Seite befinden sich die bekannten, beziehungsweise die zu bestimmenden, Elemente.	11
2.3	Die drei gefundenen Komponenten der Mischung aus der Abbildung 2.1 (b). Die Reihenfolge der Signale ist vertauscht und die Energie ist verändert.	12
2.4	Die Unsicherheiten der ICA. (a) Scatterplot zweier gemischten Signale. (b) Die unabhängigen Komponenten und (c) deren Skalierung sowie (d) eine Permutation.	13
2.5	Gemischte Signale, die nicht mit Hilfe linearer Methoden separiert werden können. (a) Eine Mischung mit mehr Komponenten als Beobachtungen und (b) eine Mischung von Signalen mit Gaußverteilung.	14
3.1	Zwei Beispiele für unkorrelierte Datensätze. (a) Eine Mischung aus zwei supergaußschen und (b) subgaußschen Signalen. In beiden Fällen ist die Korrelationsmatrix diagonal.	23
3.2	Zwei Beispiele für unabhängige Datensätze. Die Daten aus Abbildung 3.1 wurden so gedreht, dass sie unabhängig sind. Die Korrelationsmatrix ist dabei unverändert geblieben.	24
3.3	Principal Component Analysis (a) Die Verteilung der Daten wird durch den Mittelpunkt \mathbf{m}_0 und die beiden Hauptkomponenten \mathbf{v}_1 und \mathbf{v}_2 beschrieben. (b) Nach der Subtraktion des Mittelwertes und einer orthogonalen Transformation sind die Hauptkomponenten entlang der Koordinatenachsen ausgerichtet.	25

Abbildungsverzeichnis

3.4	Das Whitening-Verfahren. (a) Die Mischung zweier Signale. (b) Dekorrelation mittels der PCA nach Gleichung (3.3). (c) Die Signale nach dem Whitening nach (3.19). (d) Unabhängige Signale können nach dem Whitening mittels einer einfachen Drehung gewonnen werden.	30
3.5	Beispiel für Randverteilungen von supergaußschen Signalen. (a) Zwei gemischte Signale und die dazugehörigen Randverteilungen. (b) Zwei getrennte Signale und die dazugehörigen Randverteilungen.	32
3.6	Beispiel für Randverteilungen von subgaußschen Signalen. (a) Zwei gemischte Signale und die dazugehörigen Randverteilungen. (b) Zwei getrennte Signale und die dazugehörigen Randverteilungen.	33
3.7	Beispiel für Abhängigkeit der Kurtosis vom Projektionsvektor. (a) Streubild zweier gemischter supergaußscher Signale mit einem Projektionsvektor. (b) Die Kurtosis der projizierten Signale in Abhängigkeit des Winkels des Projektionsvektors.	36
3.8	Beispiel für Abhängigkeit der Kurtosis vom Projektionsvektor. (a) Streubild zweier gemischter subgaußscher Signale mit einem Projektionsvektor. (b) Die Kurtosis der projizierten Signale in Abhängigkeit des Winkels des Projektionsvektors.	37
3.9	Die robusten Nichtlinearitäten aus den Gleichungen (3.43) bis (3.45) und die dazugehörigen Ableitungen. $G_3(y)$ ist die Kurtosis, die zu Vergleichszwecken mitangeführt wird.	42
3.10	Beispiele für den Zusammenhang zwischen Verteilungsdichten und den Randverteilungsdichten (a) Separierbare Verteilungsdichte und (b) nicht separierbare Verteilungsdichte.	46
3.11	Zur Berechnung der KLD. (a) Verteilungsdichte, die sich aus der Multiplikation der Randverteilungen aus Abbildung 3.10 (b) ergibt. (b) Der Quotient aus Gleichung (3.57) gebildet zwischen der Verteilungsdichte aus Abbildung 3.11 (a) und 3.10 (b)	47
3.12	Beispiel für eine nichteuklidische Metrik. (a) Die kürzesten Wege zwischen zwei Punkten auf einer Kugeloberfläche sind in rot dargestellt. (b) Die kürzesten Verbindungen zwischen den gleichen Punkten bei einer euklidischen Fläche sind in blau markiert. . .	50

3.13	Zwei typische Verteilungsdichten und die dazugehörigen nichtlinearen Aktivierungsfunktionen.	53
3.14	Die generalisierte Gaußdichte für verschiedene Parameter und die dazugehörigen Aktivierungsfunktionen.	54
3.15	Schätzung des Form-Parameters β . Auf der linken Seite $M(\beta)$ und rechts die approximierte <i>generalized Gaussian ratio function</i>	57
3.16	Verteilungsdichten komplexwertiger Zufallsvariablen. (a) Verteilungsdichte nach Gleichung (3.97). (b) Verteilungsdichte nach Gleichung (3.99)	60
3.17	Zirkulär symmetrische GGD mit $\beta = 10$	62
4.1	Unterschiedliche Modelle für Raumimpulsantworten. (a) Das Freifeldmodell mit Verzögerung aufgrund der Laufzeit. (b) Das Raummodell mit vielen zusätzlichen Wiederholungen aufgrund von Reflexionen. (c) Spiegelbildmethode zur Simulation von Raumimpulsantworten.	65
4.2	Verschiedene Darstellungsformen einer Impulsantwort. (a) Eine Raumimpulsantwort im Zeitbereich. (b) Die Energie der einzelnen Samples einer Impulsantwort. (c) Betragsfrequenzgang.	66
4.3	Die Berechnung eines Spektrogramms. (a) Ein Sprachsignal. (b) Gefensterter Abschnitt. (c) Das fertige Spektrogramm.	69
4.4	Übersicht über typische Fenster. (a) In der Zeitdarstellung mit $L_\gamma = 32$. (b) Der zugehörige Frequenzgang.	72
4.5	Berechnung des Inversen Systems mit Hilfe des Direct Inverse Equalizers.	74
4.6	Reduktion eines SIMO Systems nach Gleichung (4.29)	75
4.7	Die Berechnung des Inversen mit Hilfe des Mint Theorems	76
4.8	Das allgemeine System der konvolutiven blinden Quellentrennung. Sowohl das Mischungssystem $\mathbf{H}(e^{j\omega})$ wie auch das Entmischungssystem $\mathbf{W}(e^{j\omega})$ sind MIMO Systeme.	77
4.9	Der Spezialfall des (2×2) -Systems	78
4.10	Das Entmischungssystem im Frequenzbereich mit der Korrektur der Permutation und Skalierung.	80

Abbildungsverzeichnis

4.11	Die einzelnen Schritte der konvolutiven Trennung im Frequenzbereich. (a) Zwei Aufnahmen gemischter Signale in der Zeitdarstellung. (b) Die Aufnahmen in der Zeit-Frequenz Darstellung. (c) Die Zeit-Frequenz Darstellung nach der instantanen ICA in jedem Frequenzband. (d) Die getrennten Signale im Zeitbereich.	81
4.12	Beispiele für die fehlende Korrektur des Permutations- und Skalierungsproblems. (a) Das Permutationsproblem im Zeit-Frequenzbereich. (b) Die dazugehörigen Zeitsignale sind nicht getrennt. (c) Falsche Skalierung im Frequenzbereich. (d) Die Zeitsignale sind zwar getrennt, aber stark verzerrt.	83
5.1	Die Lösung des Skalierungsproblems unter der Annahme gleicher Energie in allen Frequenzen. (a) Die Zeit-Frequenz-Darstellung eines Sprachsignals zeigt sehr große Energieunterschiede. (b) Entmischung unter der Annahme gleicher Energie in allen Frequenzbändern.	90
5.2	Die Lösung des Skalierungsproblems unter der Annahme gleicher Länge der Entmischungsvektoren. Die Energieverteilung über die Frequenzbänder ist grob erhalten, trotzdem haben, im Vergleich zu Abbildung 5.1 (a), viele Frequenzbänder eine zu kleine Energie.	91
5.3	Die Visualisierung der Bestimmung der Komponenten $x_{ij}(n)$. (a) Mischung zweier Signale mit eingezeichneten Entmischungsvektoren $\mathbf{w}_1(\omega_k)$ und $\mathbf{w}_2(\omega_k)$ nach Gleichung (5.4). (b) Getrennte Daten. (c) Die beiden Ausgangssignale ausgewählt mit der Selektionsmatrix \mathbf{S}^i . (d) Die Anwendung der inversen Nachfilter auf die einzelnen Teilsignale.	93
5.4	Die Skalierung mit inversen Nachfiltern. Die Entmischungsvektoren sind so skaliert, dass die getrennten Komponenten die gleiche Skalierung wie die Mischung haben. (a) Vektoren, die in eine ähnliche Richtung zeigen, werden verstärkt. (b) Bei nahezu senkrechten Vektoren ist eine sehr kleine Verstärkung nötig.	94
5.5	Die Skalierung unter der Annahme einer konstanten Länge der Entmischungsvektoren. (a) Bei Entmischungsvektoren, die in eine ähnliche Richtung zeigen, haben die rekonstruierten Signale eine zu kleine Energie. (b) Nur bei nahezu senkrechten Entmischungsvektoren ist die Energie der getrennten Signale erhalten.	95
5.6	Die Lösung des Skalierungsproblems mit Hilfe des MDP.	96

5.7	Projektion der Mischungs- und Entmischungsvektoren für die blinde Bestimmung der Trennleistung. Bei Einhaltung der Biorthogonalitätsbedingung sind die Signale perfekt getrennt.	98
5.8	Die Projektion der Mischungs- auf die Entmischungsvektoren bei der Verletzung der Biorthogonalitätsbedingung. Die Signale sind vollständig getrennt.	99
5.9	Beispiel für die Bestimmung der Trennleistung. (a) Die blind geschätzte Trennleistung nach Gleichung (5.22) und (b) die wahre Trennleistung.	102
5.10	Der Scatterplot der geschätzten Trennleistung gegenüber der wahren Trennleistung. Es ist eine hohe Übereinstimmung zu erkennen.	103
5.11	Beispiel für die Berechnung der Skalierungskoeffizienten mit der Max-SIR Methode. (a) Die Energieverteilung in den Frequenzbändern eines Sprachsignals. (b) Die resultierenden Skalierungskoeffizienten $c_1(\omega_k)$ nach Gleichung (5.40). Die tiefen Frequenzen werden unterdrückt, was zu einem hohen SIR aber auch starken Verzerrungen führt.	105
5.12	Beispiel für die Berechnung der Skalierungskoeffizienten mit der Max-SIR Methode. (a) Die Skalierungskoeffizienten nach Gleichung (5.42) und (b) nach Gleichung (5.43).	107
5.13	Ein Kanal des Entmischungssystems von der j -ten Mischung zu dem i -ten Ausgang, welches durch das Filter $w_{ij}(n)$ realisiert wird. Das Permutationsproblem wird in dieser Darstellung vernachlässigt.	109
5.14	Die Verkürzung der Impulsantworten. (a) Die Impulsantworten berechnet mit Hilfe des MDP. (b) Filter nach der Filterverkürzungsmethode.	114
5.15	Alternative Darstellung der Filter aus Abbildung 5.14. Gezeigt sind die Energien der einzelnen Koeffizienten im Zeitbereich. (a) Die Impulsantworten berechnet mit Hilfe des MDP. (b) Filter nach der Filterverkürzungsmethode.	115
5.16	Beispiel für die Berechnung der Skalierungsfaktoren mit Hilfe der Filterverkürzungsmethode.	116

Abbildungsverzeichnis

5.17	Fenster für die Maximierung des Betrages des Hauptmaximums. Oben das Fenster γ_d für den erwünschten Bereich und unten das komplementäre Fenster γ_u , welches den unerwünschten Bereich auswählt.	119
5.18	Fenster für die Konzentration der Energie in einem Bereich. Diese Wahl der Fenster erlaubt eine bessere Unterdrückung außerhalb des erwünschten Bereiches.	120
5.19	Fenster für Filterformung mit exponentiellen Verläufen. Diese Form führt zu Impulsantworten, die Raumimpulsantworten nachempfunden sind.	121
5.20	Doppelt exponentielle Fenster für die Filterformung. Unterschiedlichen Exponenten führen zu Impulsantworten, bei denen der Bereich vor dem Hauptmaximum stärker als der dahiner gedämpft wird. (a) Die zu formende Impulsantwort. (b) Die doppelt exponentiellen Verläufe der Fenster ausgerichtet auf das Hauptmaximum der Impulsantwort.	123
5.21	Beispiel für Filterformung. (a) Die Filter im Zeitbereich. (b) Die Energie der einzelnen Koeffizienten. Der Vergleich mit Abbildung 5.15 zeigt bessere Dämpfung der weiter entfernten Koeffizienten.	124
5.22	Beispiel für die Berechnung der Skalierungsfaktoren mit Hilfe der Filterformung.	125
6.1	Beispiel für nicht-blinde Bestimmung der Permutationen. (a) Entpermutation nach Gleichung (6.5). (b) Entpermutation durch Korrelation mit den Eingangssignalen. (c) Der Unterschied zwischen den beiden Methoden. Eine Eins markiert einen Unterschied.	131
6.2	Der zeitliche Verlauf der Einhüllenden von drei Frequenzbändern eines Sprachsignals.	133
6.3	Beispiel für die Koeffizienten $r_{pq}(\omega_k, \omega_l)$ bei zwei entpermutierten Quellen. Die blauen und grünen Bereiche weisen auf eine Permutation hin, die aber nicht vorhanden ist.	135
6.4	Koeffizienten aus Abbildung 6.3. In schwarz sind die Koeffizienten markiert, die kleiner als eins sind, und damit eine nicht vorhandene Permutation andeuten.	136
6.5	Die sequentielle Entpermutation. (a) Die Trennleistung in den Frequenzbändern. (b) Permutationen gegenüber einer nicht-blinden Methode.	137

6.6	Dyadische Sortierung. (a) Die Trennleistung in den Frequenzbändern. (b) Permutatationen gegenüber einer nicht-blinden Methode.	138
6.7	Beispiel für den Verlauf des β -Parameters der GGD in den einzelnen Frequenzbändern für zwei Signale. (a) Der Verlauf über alle Frequenzbänder. (b) Ein Ausschnitt. (c) Die Differenz der Werte.	140
6.8	Beispiel für den Verlauf des α -Parameters der GGD in den einzelnen Frequenzbändern für zwei Signale. (a) Der Verlauf über alle Frequenzbänder. (b) Ein Ausschnitt. (c) Die Differenz der Werte.	142
6.9	Der Verlauf von $\beta_{pq}(\omega_k)$ in blau nach Gleichung (6.13) und in grün der Prädiktionsfehler $\beta_{d_1}(\omega_k)$ nach Gleichung (6.23).	146
6.10	Der Verlauf von $\alpha_{pq}(\omega_k)$ in blau nach Gleichung (6.14) und in grün der Prädiktionsfehler $\alpha_{d_1}(\omega_k)$ nach Gleichung (6.26).	146
6.11	Die Verläufe der Fehlersignale sowie die geclusterten Bereiche. (a) Berechnung für $\alpha_{d_1}(\omega_{k+1})$ und $\alpha_{d_2}(\omega_{k+1})$. (b) Berechnung für $\beta_{d_1}(\omega_{k+1})$ und $\beta_{d_2}(\omega_{k+1})$. (c) Berechnung für das gemeinsame Kriterium mit $\Delta_1(\omega_{k+1})$ und $\Delta_2(\omega_{k+1})$.	148
6.12	Die Entpermutation mit der $\alpha\beta$ -Methode. (a) Die Trennleistung in den Frequenzbändern. (b) Permutatationen gegenüber einer nicht-blinden Methode.	150

Abbildungsverzeichnis

Tabellenverzeichnis

3.1	Beispiele für diverse Verteilungsdichten $p_y(y)$ und die dazugehörigen Aktivierungsfunktionen $f(y)$	52
5.1	Vergleich der Trennleistungen in dB der Max-SIR Methode, sowie der durch das Gesamtsystem (SFM _{ges}) und Entmischungssystem (SFM _{ent}) verursachten Verzerrungen.	106
5.2	Vergleich der Trennleistungen in dB mit der Filterverkürzungsmethode, sowie der durch das Gesamtsystem (SFM _{ges}) und Entmischungssystem (SFM _{ent}) verursachten Verzerrungen.	113
5.3	Vergleich der Trennleistungen in dB mit dem filter shaping Prinzip, sowie der durch das Gesamtsystem (SFM _{ges}) und Entmischungssystem (SFM _{ent}) verursachten Verzerrungen.	122
6.1	Der Zusammenhang zwischen der Länge des linearen Prädiktors L_a und dem Fehler $E\{ \beta_{d_1}(\omega_{k+1}) ^2\}$	146
6.2	Vergleich der Clustergrößen der $\alpha\beta$ -Verfahrens. Nur Cluster mit mehr als acht Frequenzbändern wurden gezählt.	147
6.3	Vergleich der Trennleistung der verschiedenen Algorithmen . . .	149

Tabellenverzeichnis

Literaturverzeichnis

- [1] R. Aichner, H. Buchner, S. Araki, and S. Makino. On-line time-domain blind source separation of nonstationary convolved signals. In *Proc. 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 987–992, Nara, Japan, April 2003.
- [2] S.-I. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10:271–276, 1998.
- [3] S.-I. Amari and J.-F. Cardoso. Blind source separation-semiparametric statistical approach. *IEEE Transactions on Signal Processing*, 45(11):2692–2700, Nov 1997.
- [4] S.-I. Amari, T.-P. Chen, and A. Cichocki. Stability analysis of adaptive blind source separation. *Neural Networks*, 10:1345–1351, 1997.
- [5] S.-I. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems*, volume 8, MIT Press, Cambridge, MA, 1996.
- [6] S.-I. Amari, A. Cichocki, and H. H. Yang. *Unsupervised Adaptive Filtering*, chapter Blind Signal Separation and Extraction. Wiley, 1999.
- [7] S.-I. Amari and S. C. Douglas. Why natural gradient? In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 1213–1216, May 12–15, 1998.
- [8] G. Arslan, B. L. Evans, and S. Kiaei. Equalization for discrete multitone transceivers to maximize bit rate. *IEEE Trans. on Signal Processing*, 49(12):3123–3135, Dec 2001.
- [9] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and deconvolution. *Neural Computation*, 7:1129–1159, 1995.
- [10] D. H. Brandwood. A complex gradient operator and its application in adaptive array theory. *IEE Proceedings F Communications, Radar and Signal Processing*, 130(1):11–16, February 1983.

- [11] H. Brehm and W. Stammer. Description and generation of spherically invariant speech-model signals. *Signal Process.*, 12(2):119–141, 1987.
- [12] H. Buchner, R. Aichner, and W. Kellermann. *Blind source separation for convolutive mixtures: A unified treatment*. Audio Signal Processing for Next-Generation Multimedia Communication Systems. Boston/Dordrecht/London: Kluwer, Feb 2004.
- [13] J.-F. Cardoso. On the stability of source separation algorithms. Proc. Workshop on Neural Networks for Signal Processing (NNSP'98):13–22, 1998.
- [14] J.-F. Cardoso. Entropic contrasts for source separation: geometry and stability. *Unsupervised adaptive filtering*, 1:139–189, 2000.
- [15] J.-F. Cardoso and B. H. Laheld. Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, 44(12):3017–3030, Dec 1996.
- [16] S. Choi, A. Cichocki, and S. Amari. Flexible independent component analysis. In T. Constantinides, S. Y. Kung, M. Niranjan, and E. Wilson, editors, *Neural Networks for Signal Processing VIII*, pages 83–92, 1998.
- [17] S. Choi., A. Cichocki, and S.-I. Amari. Flexible independent component analysis. *J. VLSI Signal Process. Syst.*, 26(1/2):25–38, 2000.
- [18] A. Cichocki and S.-I. Amari. *Adaptive Blind Signal and Image Processing*. Wiley, Chichester, 2002.
- [19] A. Cichocki, R. Thawonmas, and S.-I. Amari. Sequential blind signal extraction in order specified by stochastic properties. *Electronics Letters*, 33(1):64–65, January 1997.
- [20] A. Cichocki and R. Unbehauen. Robust neural networks with on-line learning for blind identification and blind separation of sources. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 43(11):894–906, Nov 1996.
- [21] P. Comon. Independent component analysis—a new concept? *Signal Processing*, 36(3):287–314, Apr 1994.
- [22] P. Comon. Contrasts for multichannel blind deconvolution. *Signal Processing letters*, 3(7):209–211, 1996.
- [23] N. Delfosse and P. Loubaton. Adaptive blind separation of independent sources: A deflation approach. *Signal Processing*, 45(1):59 – 83, 1995.

- [24] K. I. Diamantaras and S. Y. Kung. *Principal Component Neural Networks: Theory and Applications*. Wiley, 1996.
- [25] S. C. Douglas, H Sawada, and S. Makino. Natural gradient multichannel blind deconvolution and speech separation using causal FIR filters. *IEEE Trans. Speech and Audio Processing*, 13(1):92–104, Jan 2005.
- [26] S. Gazor and W. Zhang. Speech probability distribution. *IEEE Signal Processing Letters*, 10(7):204–207, July 2003.
- [27] M. Girolami and C. Fyfe. Negentropy and kurtosis as projection pursuit indices provide generalized ica algorithms. *Advances in Neural Information Processing Systems, NIPS '96 Workshop*, Dec 1996.
- [28] Y. Huang, J. Benesty, and J. Chen. *Acoustic MIMO Signal Processing*. Number XIV in Signals and Communication Technology. Springer, 2006.
- [29] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Adaptive and Learning Systems for Signal Processing, Communications, and Control. Wiley, 2001.
- [30] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.
- [31] A. Hyvärinen and E. Oja. Independent component analysis by general non-linear hebbian-like learning rules. *Signal Processing*, 64:301–313, 1998.
- [32] S. Ikeda and N. Murata. A method of blind separation based on temporal structure of signals. In *Proc. Int. Conf. on Neural Information Processing*, pages 737–742, 1998.
- [33] M. Z. Ikram and D. R. Morgan. Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment. In *Proc. ICASSP2000*, volume II, pages 1041–1044, 2000.
- [34] M. Z. Ikram and D. R. Morgan. Permutation inconsistency in blind speech separation: investigation and solutions. *IEEE Transactions on Speech and Audio Processing*, 13(1):1–13, Jan. 2005.
- [35] James D. Johnston. Transform coding of audio signals using perceptual noise criteria. *IEEE Journal on Selected Areas in Communication*, 6(2):314–232, Feb 1988.
- [36] I.T. Jolliffe. *Principal Component Analysis*, volume XXIX. Springer Series in Statistics, 2002.

Literaturverzeichnis

- [37] M. Kallinger. *Neue Ansätze zur Unterdrückung akustischer Echos unter Einbeziehung einer Stereo-Sprachübertragung*. PhD thesis, University of Bremen, Germany, Aachen, Germany, Jan 2007.
- [38] Markus Kallinger and Alfred Mertins. Multi-channel room impulse response shaping - a study. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, volume V, pages 101–104, Toulouse, France, May 2006.
- [39] K. D. Kammeyer and K. Kroschel. *Digitale Signalverarbeitung. Filterung und Spektralanalyse mit MATLAB-Übungen*. B.G. Teubner Verlag, 11 2002.
- [40] K. Kokkinakis and A. K. Nandi. *Multichannel Speech Separation Using Adaptive Parameterization of Source PDFs*, volume 3195 of *Lecture Notes in Computer Science*. Springer, 2004.
- [41] K. Kokkinakis and A. K. Nandi. Exponent parameter estimation for generalized gaussian probability density functions with application to speech modeling. *Signal Processing*, 85(9):1851–1858, Sept. 2005.
- [42] I. Lee, T. Kim, and T.-W. Lee. Independent vector analysis for convolutive blind speech separation. In *Blind Speech Separation*, pages 169–192. Springer Netherlands, 2007.
- [43] Hualiang Li and Tülay Adalı. Complex-valued adaptive signal processing using nonlinear functions. *EURASIP Journal on Advances in Signal Processing*, 2008:1–9, 2008.
- [44] S.G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7):674–693, Jul 1989.
- [45] Richard K. Martin, Ming Ding, Brian L. Evans, and C. Richard Johnson, Jr. Efficient channel shortening equalizer design. *EURASIP J. Appl. Signal Process.*, 2003:1279–1290, 2003.
- [46] K. Matsuoka. Minimal distortion principle for blind source separation. In *Proceedings of the 41st SICE Annual Conference*, volume 4, pages 2138–2143, 5-7 Aug. 2002.
- [47] R. Mazur and A. Mertins. Reducing reverberation effects in convolutive blind source separation. In *Proc. European Signal Processing Conference*, Florence, Italy, Sept. 2006.

- [48] R. Mazur and A. Mertins. Solving the permutation problem in convolutive blind source separation. In M. E. Davies, Ch. J. James, S. A. Abdallah, and M. D. Plumbley, editors, *Independent Component Analysis and Signal Separation*, volume 4666 of *LNCS*, pages 512–519. Springer, 2007.
- [49] R. Mazur and A. Mertins. On separation performance enhancement in convolutive blind source separation. In *Proc. Asilomar Conference on Signals, Systems, and Computers*, pages 1718–1721, Oct. 2008.
- [50] R. Mazur and A. Mertins. An approach for solving the permutation problem of convolutive blind source separation based on statistical signal models. *IEEE Trans. Audio, Speech, and Language Processing*, 17(1):117–126, Jan. 2009.
- [51] R. Mazur and A. Mertins. A method for filter shaping in convolutive blind source separation. In *Independent Component Analysis and Signal Separation (ICA2009)*, volume 5441 of *LNCS*, pages 282–289. Springer, 2009.
- [52] R. Mazur and A. Mertins. Simplified formulation of a depermutation criterion in convolutive blind source separation. In *Proc. European Signal Processing Conference*, pages 1467–1470, Glasgow, Scotland, Aug 2009.
- [53] R. Mazur and A. Mertins. Using the scaling ambiguity for filter shortening in convolutive blind source separation. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, pages 1709–1712, Taipei, Taiwan, April 2009.
- [54] T. Mei, A. Mertins, and M. Kallinger. Room impulse response reshaping/shortening based on least mean squares optimization with infinity norm constraint. In *Proc. 16th International Conference on Digital Signal Processing*, pages 1–6, July 5–7, 2009.
- [55] T. Mei, J. Xi, F. Yin, A. Mertins, and J. F. Chicharo. Blind source separation based on time-domain optimizations of a frequency-domain independence criterion. *IEEE Trans. Audio, Speech, and Language Processing*, 14(6):2075–2085, Nov. 2006.
- [56] P.J.W. Melsa, R.C. Younce, and C.E. Rohrs. Impulse response shortening for discrete multitone transceivers. *IEEE Trans. on Communications*, 44(12):1662–1672, Dec. 1996.
- [57] A. Mertins, Tiemin Mei, and M. Kallinger. Room impulse response shortening/reshaping with infinity- and p -norm optimization. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(2):249–259, February 2010.

Literaturverzeichnis

- [58] M. Miyoshi and Y. Kaneda. Inverse filtering of room acoustics. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(2):145–152, February 1988.
- [59] L. Molgedey and H. G. Schuster. Separation of a mixture of independent signals using time-delayed correlations. *Physical Review Letters*, 72(23):3634–3637, 1994.
- [60] J.A.D. Molina, G.G. Farias, and R.M.R. Dagnino. A practical procedure to estimate the shape parameter in the generalized gaussian distribution. http://www.cimat.mx/reportes/enlinea/I-01-18_eng.pdf.
- [61] R. Mukai, H. Sawada, S. Araki, and S. Makino. Frequency-domain blind source separation of many speech signals using near-field and far-field models. In *EURASIP Journal on Applied Signal Processing*, 2006. Article ID 83683 13 Pages.
- [62] J.-P. Nadal, E. Korutcheva, and F. Aires. Blind source separation in the presence of weak sources. *Neural Networks*, 13(6):589–596, 2000.
- [63] S. T. Neely and J. B. Allen. Invertibility of a room impulse response. *J. Acoustical Society of America*, 68:165–169, July 1979.
- [64] E Oja. Principal components, minor components, and linear neural networks. *Neural Networks*, 5:927 – 935, 1992.
- [65] Alan V. Oppenheim, Ronald W. Schaffer, and John R. Buck. *Discrete-time signal processing (2nd ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1999.
- [66] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw Hill, 1984.
- [67] L. Parra and C. Spence. Convolutional blind source separation of non-stationary sources. *IEEE Trans. Speech Audio Processing*, 8(3):320–327, May 2000.
- [68] Karl Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559 – 572, 1901.
- [69] D.-T. Pham, Ch. Servière, and H. Boumaraf. Blind separation of convolutional audio mixtures using nonstationarity. In *Proceedings of the 4th International Conference on Independent Component Analysis and Blind Signal Separation (ICA '03)*, pages 975–980, Nara, Japan, April 2003.

- [70] B. D. Radlovic and R. A. Kennedy. Nonminimum-phase equalization and its subjective importance in room acoustics. *IEEE Transactions on Speech and Audio Processing*, 8(6):728–737, November 2000.
- [71] K. Rahbar and J. P. Reilly. Blind source separation of convolved sources by joint approximate diagonalization of cross-spectral density matrices. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, volume 5, pages 2745–2748, 7-11 May 2001.
- [72] K. Rahbar and J. P. Reilly. A frequency domain method for blind source separation of convolutive audio mixtures. *IEEE Trans. Speech and Audio Processing*, 13(5):832–844, Sept. 2005.
- [73] H. Sawada, R. Mukai, S. Araki, and S. Makino. A robust and precise method for solving the permutation problem of frequency-domain blind source separation. *IEEE Trans. Speech and Audio Processing*, 12(5):530–538, Sept. 2004.
- [74] D. Schobben, K. Torkkola, and P. Smaragdis. Evaluation of blind signal separation methods. In *Proc. Int. Workshop Independent Component Analysis and Blind Signal Separation*, Aussois, France, Jan. 1999.
- [75] R. Thawonmas, A. Cichocki, and S.-I. Amari. A cascade neural network for blind signal extraction without spurious equilibria. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E81-A(9):1833–1846, 1998.
- [76] <http://www.kecl.ntt.co.jp/icl/signal/sawada/demo/bss2to4/index.html>.
- [77] A. van den Bos. Complex gradient and hessian. *IEEE Proceedings - Vision, Image and Signal Processing*, 141(6):380–383, December 1994.
- [78] M. K. Varanasi and B. Aazhang. Parametric generalized gaussian density estimation. *Acoustical Society of America Journal*, 86(4):1404–1415, Oct 1989.
- [79] E. Vincent. Complex nonconvex lp norm minimization for underdetermined source separation. In M. E. Davies, Ch. J. James, S. A. Abdallah, and M. D. Plumbley, editors, *Independent Component Analysis and Signal Separation*, volume 4666 of *LNCS*, pages 430–437. Springer, 2007.
- [80] W. Wang, J. A. Chambers, and S. Sanei. A novel hybrid approach to the permutation problem of frequency domain blind source separation. In

Literaturverzeichnis

Lecture Notes in Computer Science, volume 3195, pages 532–539. Springer, 2004.

- [81] W. Wirtinger. Zur formalen theorie der funktionen von mehr komplexen veränderlichen. *Mathematische Annalen*, 97(1):357, December 1926.
- [82] H.-C. Wu and J. C. Principe. Generalized anti-hebbian learning for source separation. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP '99*, volume 2, pages 1073–1076, March 15–19, 1999.