

# Verbesserung der Quellentrennung bei mehrkanaligen Audiosignalmischungen mittels neuartiger Ansätze zur Permutationskorrektur und zur Zeit-Frequenz-Maskierung

vorgelegt von  
Dipl.-Ing. Eugen Hoffmann  
aus Berlin

von der Fakultät IV - Elektrotechnik und Informatik  
der Technischen Universität Berlin  
zur Erlangung des akademischen Grades  
Doktor der Ingenieurwissenschaften  
- Dr.-Ing. -  
genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr.-Ing. S. Möller

Gutachter: Prof. Dr.-Ing. R. Orglmeister

Gutachter: Prof. Dr.-Ing. B. Meffert

Tag der wissenschaftlichen Aussprache: 20. September 2013

Berlin 2013

D 83



# Vorwort

Die vorliegende Dissertation entstand während meiner Zeit als wissenschaftlicher Mitarbeiter am Fachgebiet Elektronik und medizinische Signalverarbeitung der Technischen Universität Berlin. Mein herzlicher Dank gilt allen Mitarbeitern des Instituts, die zum Gelingen dieser Arbeit beigetragen haben.

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

Frau Prof. Dr.-Ing. B. Meffert danke ich für die Übernahme des Koreferats.

Herrn Paul Kiesow danke ich für die Übernahme der Korrekturen. Weiterhin möchte ich B.G. Yun, L.I. Grigoriadi, W.I. Asimzew, N.P. Bespalowa, G.M. Balaschowa, E.K. Hottmann, O.J. Hottmann, E. Jöckers, R. und S. Nowitzkij, W. Hoffmann, K. Milner, A. Reisig, Y. Vatis, M. Balaev, R. Sharipov, A. Fomin, V. Mysotschkin, A. Ermolin, A. Emelyanov sowie allen meinen lieben Freunden für ihre Hilfe, Unterstützung, Ausdauer, Geduld und ihren Glauben in meine Kräfte danken, womit sie mir stets zur Seite standen.

Schließlich gilt herzlicher Dank auch meinen Eltern, die mich während der gesamten Zeit tatkräftig unterstützt haben.

Berlin, 27. September 2013

Eugen Hoffmann

# Abstract

Blinde Quellentrennung (Blind Source Separation, BSS) ist ein Verfahren zur Rekonstruktion der Quellsignale aus beobachteten Mischungen, wobei weitere Informationen über die Quellen, Quellenkonstellation oder Mischungsart nicht vorliegen. Mögliche Einsatzgebiete der Blinden Quellentrennung könnten beispielsweise rauschrobuste Spracherkennung, sprachgesteuerte Telekommunikationssysteme und Hörhilfen sein. Der Einsatz von Verfahren der Blinden Quellentrennung in realen Umgebungen ist jedoch mit gewissen Problemen verbunden. Zum einen tritt bei der Entmischung ein so genanntes Permutationsproblem auf, das durch die Eigenschaften der BSS-Algorithmen entsteht und das es, um korrekte Ergebnisse zu bekommen, zu lösen gilt. Zum anderen kommt es dazu, dass die Ausgangssignale durch bleibende Übersprecher gestört sind. Die Idee der Zeit-Frequenz-Maskierung ist es, die vorhandenen Störgeräusche durch Verwendung einer geeigneten Maske zu minimieren. So werden unterschiedliche Zeit-Frequenz-Maskierungsverfahren immer öfter als eine Nachverarbeitungsstufe bei BSS bzw. Beamforming eingesetzt.

Das Ziel der Arbeit sind die Entwicklung und der Einsatz neuer Algorithmen zur Permutationskorrektur sowie von Zeit-Frequenz-Maskierungsverfahren und der Vergleich der neuen Algorithmen mit den bereits existierenden. Zur Lösung des Problems werden die statistischen, räumlichen und informationstheoretischen Eigenschaften der vorliegenden Signale untersucht. Weiterhin werden die bereits existierenden Verfahren zur einkanaligen Störgeräuschreduktion auf Erweiterbarkeit und Anwendbarkeit in einem mehrkanaligen Szenario bzw. für die Verbesserung der bereits existierenden Zeit-Frequenz-Maskierungsverfahren geprüft. Zur Bewertung und zum Vergleich der resultierenden Algorithmen dienen neben den objektiven Bewertungsmaßen (wie Signal-to-Interference-Ratio und Signal-to-Distortion-Ratio) auch subjektive auditive Bewertungen in Form von Hörtests.

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>ix</b>
<b>Tabellenverzeichnis</b>	<b>xii</b>
<b>1. Einleitung</b>	<b>1</b>
<b>2. Mathematische Grundlagen</b>	<b>5</b>
2.1. Grundlagen der Statistik . . . . .	5
2.1.1. Zufallsvariablen . . . . .	5
2.1.2. Momente . . . . .	6
2.1.3. Kumulanten . . . . .	7
2.1.4. Verteilungsdichten . . . . .	7
2.1.4.1. Gaußverteilung . . . . .	8
2.1.4.2. Generalisierte Gaußverteilung . . . . .	8
2.1.4.3. Exponentialverteilung . . . . .	10
2.1.4.4. Rayleigh-Verteilung . . . . .	10
2.1.4.5. Weibull-Verteilung . . . . .	11
2.1.4.6. Rice-Verteilung . . . . .	11
2.2. Grundlagen der Informationstheorie . . . . .	12
2.2.1. Entropie . . . . .	13
2.2.1.1. Entropiedefinition nach Shannon . . . . .	13
2.2.1.2. Entropiedefinition nach Rényi . . . . .	14
2.2.1.3. Entropiedefinition nach Tsallis . . . . .	15
2.2.2. Generalisierte Rényi-Divergenz . . . . .	17
2.2.3. Generalisierte Tsallis-Divergenz . . . . .	18
2.2.4. Gemeinsame Information . . . . .	19
2.2.5. Jensen-Rényi-Divergenz . . . . .	19
2.2.6. Jensen-Tsallis-q-Divergenz . . . . .	21
2.2.7. Negentropie . . . . .	21
2.3. Kernfunktionen und ihre Anwendung . . . . .	22
2.3.1. Kernel-Trick . . . . .	22

2.3.2. Korrentropie . . . . .	23
<b>3. Methoden der blinden Quellentrennung</b>	<b>27</b>
3.1. Lineares Mischungsmodell . . . . .	27
3.2. Statistische Unabhängigkeit . . . . .	29
3.3. Korrelation . . . . .	29
3.4. Algorithmen der linearen instantanen ICA . . . . .	31
3.4.1. Infomax . . . . .	31
3.4.2. FastICA . . . . .	33
3.4.3. Verfahren auf der Basis von Statistik höherer Ordnung . . . . .	36
3.4.4. Verfahren auf der Basis von Statistik zweiter Ordnung . . . . .	38
3.4.5. Unzulänglichkeiten der linearen blinden Quellentrennung . . . . .	40
<b>4. Blinde Quellentrennung konvolutiver Mischungen</b>	<b>43</b>
4.1. Konvolutives Mischungsmodell . . . . .	43
4.2. Überblick über die Verfahren zur Permutationskorrektur . . . . .	45
4.2.1. Permutationskorrekturen basierend auf der Konsistenz der Spektren . . . . .	46
4.2.2. Permutationskorrekturen basierend auf der Konsistenz der Filterkoeffizienten . . . . .	48
4.3. Permutationskorrekturen basierend auf den informationstheoretischen Abstandsmaßen . . . . .	50
4.3.1. Allgemeine Betrachtungen . . . . .	51
4.3.2. Definition der neuen Abstandsmaße . . . . .	55
4.3.3. Algorithmen zu Permutationskorrekturen . . . . .	59
4.4. Permutationskorrekturen basierend auf den Kernfunktionen . . . . .	60
<b>5. Verfahren zur Zeit-Frequenz-Maskierung und Spektralen Subtraktion</b>	<b>65</b>
5.1. Überblick über die Verfahren der Zeit-Frequenz-Maskierung . . . . .	65
5.1.1. DUET-Algorithmus . . . . .	69
5.1.2. Amplitudenbasierte Maskierung . . . . .	70
5.1.3. Phasenbasierte Maskierung . . . . .	72
5.2. Interferenzen basierte Maskierung . . . . .	73
5.2.1. Extraktion einer Maskenfunktion aus den approximierten Interferenzen . . . . .	73
5.2.2. Bestimmung der Schwellenwerte . . . . .	76
5.3. Verfahren der einkanaligen Störgeräuschreduktion . . . . .	78
5.3.1. Grundlagen der Störgeräuschreduktion . . . . .	79

5.3.2.	Verfahren der Spektralen Subtraktion . . . . .	81
5.3.2.1.	Wiener-Filter . . . . .	81
5.3.2.2.	Maximum-Likelihood-Subtraktion . . . . .	83
5.3.2.3.	Ephraim-Malah-Filter . . . . .	86
5.3.3.	Schätzung des A-Priori SNR . . . . .	89
5.3.4.	Wahrscheinlichkeit für Sprachaktivität . . . . .	90
5.3.4.1.	Berechnung der Sprachwahrscheinlichkeit nach McAulay-Mallpass . . . . .	93
5.3.4.2.	Berechnung der Sprachwahrscheinlichkeit nach I. Cohen . . . . .	95
5.4.	Anwendung der Methoden der Spektralen Subtraktion im mehrkana- ligen Szenario . . . . .	95
5.4.1.	Berechnung der Sprachwahrscheinlichkeit . . . . .	96
5.4.2.	Mehrkanalige Störgeräuschunterdrückung unter Berücksichti- gung der Sprachwahrscheinlichkeiten . . . . .	97
<b>6. Bewertungsmethoden</b>		<b>101</b>
6.1.	Bewertungsmethoden im Vergleich . . . . .	101
6.2.	Die objektiven Bewertungsmethoden . . . . .	102
6.2.1.	Signal-zu-Interferenz-Verhältnis . . . . .	103
6.2.2.	Signal-zu-Verzerrung-Verhältnis . . . . .	104
6.3.	Die subjektiven Bewertungsmethoden . . . . .	105
<b>7. Ergebnisse und Diskussion</b>		<b>107</b>
7.1.	Quelldatensätze und Meßbedingungen . . . . .	107
7.2.	Ergebnisse der Permutationskorrekturen . . . . .	109
7.2.1.	Ergebnisse der Permutationskorrekturen mit informations- theoretischen Abstandsmaßen . . . . .	111
7.2.2.	Ergebnisse der Permutationskorrekturen mit Kernfunktionen . . . . .	114
7.2.3.	Vergleich der Algorithmen zur Permutationskorrektur . . . . .	117
7.3.	Ergebnisse der Zeit-Frequenz-Maskierung . . . . .	119
7.3.1.	Ergebnisse der „klassischen“ Zeit-Frequenz-Maskierung ohne Nachfilterung . . . . .	119
7.3.2.	Ergebnisse der mehrkanaligen Störgeräuschunterdrückung . . . . .	123
7.3.3.	Subjektive Bewertung der Algorithmen zur Zeit-Frequenz- Maskierung . . . . .	127
<b>8. Zusammenfassung und Ausblick</b>		<b>133</b>

<b>A. Ausdrücke für die Shannon-, Rényi- und Tsallis-Entropien bei Vorliegen der Weibull- und der generalisierten Gauß-Verteilungen</b>	<b>137</b>
A.1. Entropien der Weibull-Verteilung . . . . .	137
A.2. Entropien der Generalisierten Gauß-Verteilung . . . . .	139
<b>Literaturverzeichnis</b>	<b>141</b>

# Abbildungsverzeichnis

2.1. Abhängigkeit der Generalisierten Gaußverteilung vom Parameter $\beta$ (für alle Kurven wurde $a = 1$ gewählt). . . . .	10
2.2. Abhängigkeit der Weibull-Verteilung vom Parameter $k$ (für alle Kurven wurde $\lambda = 1$ gewählt). . . . .	12
2.3. Abhängigkeit der Rice-Verteilung vom Parameter $\nu$ (für alle Kurven wurde $\sigma = 1$ gewählt). . . . .	13
4.1. Datenfluss bei der frequenzbasierten Entmischung konvolutiver Mischungen. . . . .	45
4.2. Dichtemodelle der Sprachsignalverteilungen unter der Annahme der Rayleigh- (schwarze Kurve) und der Weibull-Verteilungen (gestrichelte Kurve) und gemessene Histogramme der Spektralkomponenten eines Sprachsignals im Frequenzband um a) 1000 Hz, b) 2000 Hz, c) 3000 Hz und d) 4000 Hz. Das Sprachsignal wurde in einem Raum mit einer Reverberationszeit $T_R = 159$ ms aufgenommen. Für den Vergleich wurden 10 Sekunden der Aufnahme verwendet. . . . .	53
5.1. Übersicht über die Verfahren der Zeit-Frequenz-Maskierung. . . . .	67
5.2. Datenfluss bei der frequenzbasierten Entmischung konvolutiver Mischungen mit Zeit-Frequenz-Maskierung. . . . .	68
5.3. Zweidimensionales Histogramm aufgespannt für alle Paare $\langle \tilde{a}(\Omega, \tau), \tilde{\delta}(\Omega, \tau) \rangle$ . Das Histogramm wurde für die Aufnahme aus einem realen nicht reverberationsfreien Raum (Reverberationszeit $T_R = 130$ ms) mit zwei Mikrofonen und vier Sprechern berechnet. . . . .	71
5.4. Berechnung der Winkel $\theta_i(\Omega, \tau)$ im Whitening-Raum für einen Fall von $N = 2$ . Die Winkel werden zwischen den Vektoren der geschätzten Mischungsmatrix $\mathbf{O}(\Omega)$ und den dekorellierten Signalen $\mathbf{Z}(\Omega, \tau)$ berechnet. Die Vektoren $\mathbf{U}(\Omega)\mathbf{a}_1(\Omega)$ und $\mathbf{U}(\Omega)\mathbf{a}_2(\Omega)$ verdeutlichen die Lage der tatsächlichen Basisvektoren der Mischungsmatrix. . . . .	73

5.5.	Berechnung der Interferenzen für einen $2 \times 2$ Fall. Die Signal- mischung wurde in einem Raum mit einer Reverberationszeit $T_R = 159$ ms aufgenommen. Die Abbildungen 1) und 2) stellen die unteren 200 Frequenzbänder der Spektrogramme der Quellsignale dar 3) und 4) die entsprechenden ICA-Ausgänge nach der Permutationskorrek- tur, 5) die Signalaktivität $U_1(\Omega, \tau, R_\Omega, R_\tau)$ berechnet nach der Gleich- ung 5.21. 6) die Signalaktivität $U_2(\Omega, \tau, R_\Omega, R_\tau)$ . Für den $2 \times 2$ Fall ist $U_1(\Omega, \tau, R_\Omega, R_\tau) = N_2(\Omega, \tau, R_\Omega, R_\tau)$ und $U_2(\Omega, \tau, R_\Omega, R_\tau) =$ $N_1(\Omega, \tau, R_\Omega, R_\tau)$ nach der Gleichung 5.22. . . . .	75
5.6.	Gegenüberstellung der wahren Verteilungsdichte einer Nutzsinalak- tivität von einem durch ICA getrennten Signal (Gl. 5.21) und einer Rayleigh-Verteilung (schwarze Kurve). Das Sprachsignal wurde in ei- nem Raum mit einer Reverberationszeit $T_R = 159$ ms aufgenommen. Für den Vergleich wurden 10 Sekunden der Aufnahme verwendet. . .	77
5.7.	Maskierte Signale für einen $2 \times 2$ -Fall. Die Signale wurden auf der gleichen Basis aufgenommen wie in Abbildung 5.5. Die Abbildungen 1) und 2) stellen die unteren 200 Frequenzbänder der Spektrogramme von Signalen dar, die mit Hilfe der Maske aus der Gleichung 5.26 berechnet wurden. Die Berechnung der Schwellenwerte $\lambda_u$ und $\lambda_n$ erfolgte nach der Gleichung 5.32 mit $P_\lambda = 0,5$ für beide Schwellenwerte.	78
5.8.	Datenfluss bei der Spektralen Subtraktion . . . . .	80
5.9.	Datenfluss bei der Spektralen Subtraktion unter Berücksichtigung der Wahrscheinlichkeit für Sprachaktivität . . . . .	93
7.1.	Schematische Darstellung der Messanordnung. Das Quellsignal $i$ , $i \in$ $1 \dots N$ , trifft am linearen Mikrofonarray mit dem Einfallswinkel $\theta_i$ aus der Entfernung $L_i$ ein und wird vom Mikrofon $j$ , $j \in 1 \dots M$ , aufgenommen. $d$ ist der Mikrofonabstand im Array. . . . .	111
7.2.	Zusammenstellung der besten Ergebnisse im Sinne von $\Delta$ SIR (schwar- zer Balken) und SDR (weißer Balken) von Algorithmen zur Permuta- tionskorrektur basierend auf informationstheoretischen Abstandsma- ßen angewandt auf 1) Beträge der Amplituden der zu korrigierenden Frequenzbänder unter der Annahme der Weibull-Verteilung (siehe Al- gorithmus 2) und 2) logarithmierten Beträge der Amplituden unter der Annahme der GGD (siehe Algorithmus 3). Die x-Achse bezeich- net die Nummer des Distanzmaßes, so wie sie in der Tabelle 7.6 zu finden ist. Die y-Achse zeigt die Anzahl der besten Ergebnisse über alle Experimente bzw. Anzahl der Fälle, in denen mit dem jeweiligen Distanzmaß das beste Ergebnis erzielt worden ist. . . . .	113

7.3.	Zusammenstellung der besten Ergebnisse im Sinne von $\Delta$ SIR (schwarzer Balken) und SDR (weißer Balken) von Algorithmen zur Permutationskorrektur basierend auf Korrentropie angewandt auf 1) Beträge der Amplituden der zu korrigierenden Frequenzbänder und 2) logarithmierten Beträgen der Amplituden (siehe Algorithmus 4). Die x-Achse bezeichnet die Nummer der Kernfunktion, so wie sie in der Tabelle 7.8 zu finden ist. Die y-Achse zeigt die Anzahl der besten Ergebnisse über alle Experimente. . . . .	116
7.4.	Histogramm der besten Ergebnisse im Sinne von $\Delta$ SIR (schwarzer Balken) und SDR (weißer Balken) von Algorithmen zur Permutationskorrektur. Die x-Achse bezeichnet die Nummer vom Algorithmus, so wie sie in der Tabelle 7.10 zu finden ist. Die y-Achse zeigt die Anzahl der besten Ergebnisse über alle Experimente. . . . .	119
7.5.	Die Abhängigkeit der mittleren $\Delta$ SIR ( $\circ$ ) und SDR-Werte ( $\Delta$ ) von $G_{min}$ für den ML-Algorithmus mit $P(H_1^{(i)}(\Omega, \tau)   \mathbf{Y}(\Omega, \tau)) = M_i^{Int}(\Omega, \tau)$ für den Datensatz Nr. 2 aus der Tabelle 7.1 mit FastICA. . . . .	127
7.6.	Ergebnisse der subjektiven Bewertungen der Sprachverständlichkeit nach der Zeit-Frequenz-Maskierung (Tabelle 7.17) gemittelt über 22 Testpersonen für zwei Testsignale. Größere Werte in den Bewertungen entsprechen einer besseren Sprachverständlichkeit. . . . .	129
7.7.	Ergebnisse der subjektiven Bewertungen der Unterdrückung der Störsprecher durch die Zeit-Frequenz-Maskierung (Tabelle 7.17) gemittelt über 22 Testpersonen für zwei Testsignale. Größere Werte in den Bewertungen entsprechen einer besseren Unterdrückung der Störsprecher durch die Algorithmen. . . . .	130
7.8.	Ergebnisse der subjektiven Bewertungen der Verzerrungsfreiheit nach der Zeit-Frequenz-Maskierung (Tabelle 7.17) gemittelt über 22 Testpersonen für zwei Testsignale. Kleinere Werte in den Bewertungen entsprechen einer größeren Verzerrung, die im Signal durch die Algorithmen verursacht wurde. . . . .	131



# Tabellenverzeichnis

2.1. Einige Kernfunktionen im Überblick [Sil86, SS01]. . . . .	24
5.1. Übertragungsfunktionen der einzelnen Algorithmen zur Spektralen Subtraktion. . . . .	89
7.1. Mischungsparameter. . . . .	108
7.2. Mischungsparameter. . . . .	109
7.3. Mischungsparameter. . . . .	109
7.4. Mischungsparameter. . . . .	110
7.5. Mischungsparameter. . . . .	110
7.6. Verwendete Abstandsmaße. . . . .	114
7.7. Mittelung über die berechneten Ergebnisse von $\Delta$ SIR und SDR- Werten für jedes Distanzmaß (Tabelle 7.6). $M_i$ steht für den Mit- telwert von $\Delta$ SIR/SDR-Werten berechnet über alle Mischungen von $i$ Signalen bzw. Sprechern (Tabelle 7.1-7.5). . . . .	115
7.8. Verwendete Kernfunktionen. . . . .	115
7.9. Mittelung über die berechneten Ergebnisse von $\Delta$ SIR und SDR- Werten für jede Kernfunktion (Tabelle 7.8). $M_i$ steht für den Mit- telwert von $\Delta$ SIR/SDR-Werten berechnet über alle Mischungen von $i$ Signalen bzw. Sprechern (Tabelle 7.1-7.5). . . . .	117
7.10. Verwendete Verfahren zur Permutationskorrektur. . . . .	118
7.11. Mittelung über die berechneten Ergebnisse von $\Delta$ SIR und SDR- Werten für jedes Distanzmaß (Tabelle 7.10). $M_i$ steht für den Mit- telwert der $\Delta$ SIR/SDR-Werte berechnet über alle Mischungen von $i$ Signalen bzw. Sprechern (Tabelle 7.1-7.5). . . . .	120
7.12. Verwendete Algorithmen zur Zeit-Frequenz-Maskierung. . . . .	121
7.13. Ergebnisse der Zeit-Frequenz-Maskierung für die Algorithmen aus der Tabelle 7.12 in Form von mittleren $\Delta$ SIR/SDR-Werten [dB] über alle Datensätze bzw. Mischungen (Tabellen 7.1-7.5). . . . .	121

7.14. Verwendete Methoden zur Bestimmung der Sprachwahrscheinlichkeiten. Für die Berechnung der Sprachwahrscheinlichkeit mit Amplituden-, Phasen- und Interferenzbasierten Masken wurden die selben Parameter benutzt wie in der Tabelle 7.12. . . . . 123

7.15. Ergebnisse der mehrkanaligen Störgeräuschunterdrückung in Form mittlerer  $\Delta$ SIR/SDR-Werten [dB] für die Arten zur Bestimmung von Sprachwahrscheinlichkeiten aus der Tabelle 7.14 für den Maximum-Likelihood-Algorithmus (Gleichung 5.75) über alle Mischungen (Tabellen 7.1-7.5). Für die Berechnung von Rauschleistung und der Schätzung der a priori und a posteriori SNRs wurden  $\alpha_D = 0,5$  (Gleichung 5.120) und  $\alpha = 0,95$  (Gleichung 5.121) benutzt. Der Wert von  $G_{\min}$  in der Gleichung 5.101 wurde auf 0,05 gesetzt. Je kleiner  $G_{\min}$  gewählt wird, desto größer ist das SIR aber auch die Verzerrung des Signals. . . . . 125

7.16. Ergebnisse der mehrkanaligen Störgeräuschunterdrückung in Form mittlerer  $\Delta$ SIR/SDR-Werten [dB] für die Arten zur Bestimmung von Sprachwahrscheinlichkeiten aus der Tabelle 7.14 für den MMSE-Algorithmus (Gleichung 5.86) über alle Mischungen (Tabellen 7.1-7.5). Für die Berechnung von Rauschleistung und der Schätzung der a priori und a posteriori SNRs wurden  $\alpha_D = 0,5$  (Gleichung 5.120) und  $\alpha = 0,95$  (Gleichung 5.121) benutzt. Der Wert von  $G_{\min}$  in der Gleichung 5.101 wurde auf 0,05 gesetzt. Je kleiner  $G_{\min}$  gewählt wird, desto größer ist das SIR aber auch die Verzerrung des Signals. . . . . 126

7.17. Die in den Hörtests verglichenen Algorithmen zur Nachverarbeitung der Sprachsignale nach der blinden Quellentrennung. . . . . 128

7.18.  $\Delta$ SIR/SDR-Werte für die bei den Hörtests verwendeten Signale sowie die Mittelwerte der subjektiven Bewertungen für die drei getesteten Signaleigenschaften über alle getesteten Algorithmen. . . . . 130

# Notation

Bezeichnung	Erläuterung
<b>A</b>	Mischungsmatrix
$A(\Omega, \tau)$	Amplitudenverlauf des Nutzsymbols
Cum	Kumulante
$d$	Mikrofonabstand
$D(f(x)  q(x))$	informationstheoretische Distanz zwischen Verteilungsdichten $f(x)$ und $q(x)$
$D(\Omega, \tau)$	Amplitudenverlauf des Störsymbols
$E[\cdot]$	Erwartungswert
$f_X(x)$	Verteilungsdichte der Zufallsvariable $X$
$G(\Omega, \tau)$	Gainfunktion der Spektralen Subtraktion
$H(\cdot)$	Shannon-Entropie
$H_\alpha^R(\cdot)$	Renyi-Entropie der Ordnung $\alpha$
$H_q^T(\cdot)$	Tsallis-Entropie der Ordnung $q$
<b>I</b>	Einheitsmatrix
$I(\mathbf{X}_1, \dots, \mathbf{X}_N)$	gemeinsame Information
$\Im\{\cdot\}$	Imaginärteil
<b>J</b>	Jacobi-Matrix
$J(\cdot)$	Kostenfunktion
$M_i(\Omega, \tau)$	Maskierungsfunktion für das Signal $i$ im Frequenzbereich
$N_{FFT}$	Anzahl der Frequenzbänder
$\mathcal{N}$	Gaußverteilung
<b>O</b>	orthogonale Matrix
$P(A)$	Wahrscheinlichkeit für Ereignis $A$
<b>P</b>	Permutationsmatrix
$\Re\{\cdot\}$	Realteil
$s_i(t)$	Quellsymbol $i$ im Zeitbereich
$S_i(\Omega, \tau)$	Quellsymbol $i$ im Frequenzbereich
$\mathbf{T}(\cdot)$	linearer Operator
<b>U</b>	Whiteningmatrix
<b>W</b>	Entmischungsmatrix

$x_i(t)$	Sensorsignal $i$ im Zeitbereich
$X_i(\Omega, \tau)$	Sensorsignal $i$ im Frequenzbereich
$y_i(t)$	Ausgangssignal $i$ im Zeitbereich
$Y_i(\Omega, \tau)$	Ausgangssignal $i$ im Frequenzbereich
$\gamma(\Omega, \tau)$	A-Posteriori-SNR
$\kappa(\cdot, \cdot)$	Kernfunktion
$\kappa_X(x)$	Kurtosis von $X$
$\Lambda$	Diagonalmatrix
$\mu$	Mittelwert
$\xi(\Omega, \tau)$	A-Priori-SNR
$\rho_{X,Y}$	Korrelation zwischen Variablen $X$ und $Y$
$\sigma$	Varianz
$\Sigma$	Kovarianzmatrix
$\tau$	Zeitindex im Frequenzbereich
$\phi(\Omega, \tau)$	Phasenverlauf des Nutzsignals
$\Omega$	Nummer des Frequenzbins

---

## Abkürzungen

<b>Bezeichnung</b>	<b>Erläuterung</b>
AKF	Autokorrelationsfunktion
BSS	Blind Source Separation
CASA	Computational Auditory Scene Analysis
DFT	Diskrete Fouriertransformation
DOA	Direction of Arrival
FFT	Fast Fourier Transformation
GGD	Generalized Gaussian Distribution
ICA	Independent Component Analysis
MMSE	Minimum Mean Squared Error
PCA	Principal Component Analysis
SIR	Signal to Interference Ratio
SNR	Signal to Noise Ratio
SDR	Signal to Distortion Ratio

STFT	Short Time Fourier Transform
VAD	Voice Activity Detection



# 1. Einleitung

Mit der rasanten technologischen Entwicklung der letzten Jahre auf dem Gebiet der statistischen Signalverarbeitung einerseits und der enormen Steigerung der Rechenkapazitäten andererseits wächst die Anzahl der Anwendungen der Methoden der mehrkanaligen Signalverarbeitung. Bei einer großen Klasse der Anwendungen handelt es sich um das Problem der Rekonstruktion einzelner Quellsignale aus einer Signalmischung. Als Beispiele kann man an der Stelle folgende Einsatzgebiete nennen:

- Radioastronomie [Tre02],
- rausch-robuste Spracherkennung [Kol07],
- sprachgesteuerte Telekommunikationssysteme [Bau05],
- Hörhilfen [Dör98],
- medizinische Datenanalyse und Erfassung [Köh99] etc..

In einer Situation, in welcher eine Rekonstruktion einzelner Quellsignale erwünscht ist, ohne jedoch Vorwissen über die Mischungsart oder ohne weitere konkrete Informationen über den Verlauf der Quellsignale zu haben, handelt es sich um das so genannte Problem der *Blinden Quellentrennung* (engl. *Blind Source Separation, BSS*). Ein populäres Beispiel hierfür wäre das so genannte *Cocktail Party Problem*, bei dem es darum geht mehrere Sprachsignale, die in einem Raum zur selben Zeit auftauchten und durch mehrere Mikrophone gleichzeitig aufgenommen wurden, zu rekonstruieren. Blinde Quellentrennung ist also eine Technik, die es ermöglicht, aus einer mehrkanaligen Überlagerung die Quellsignale wiederherzustellen.

Die Lösung des Problems der blinden Quellentrennung beruht öfters auf dem Einsatz und der Analyse der Statistiken höherer Ordnung bzw. der (meistens physikalisch begründeten) Suche nach statistisch unabhängigen Komponenten der Mischsignale (engl. *Independent Component Analysis, ICA*). Dabei werden idealerweise die vorhandenen Signale in ihre ursprünglichen Komponenten zerlegt. Die vorliegende Arbeit widmet sich dem Problem der blinden Quellentrennung von Sprachsignalen unter natürlichen Bedingungen. D.h. die zu betrachtende Signalmischung ist durch eine

Faltung des Vektors der Quellsignale mit einer unbekanntem Filtermatrix entstanden. Die Arbeit zielt dabei auf eine Verbesserung der Sprachqualität und Sprachverständlichkeit in gestörten Umgebungen und in Umgebungen, in denen mehrere Sprecher gleichzeitig aktiv sind, ab.

Der Einsatz der ICA-Methoden in der oben beschriebenen Situation ist mit weiteren Problemen behaftet, wie z.B. Permutations- und Skalierungsproblemen, die unter Umständen nicht eindeutig gelöst werden können.

Ein weiteres Problem stellen die bleibenden Interferenzen bzw. Störsignale dar, d.h. dass es trotz Verwendung der ICA-Methoden nicht möglich ist, die Störsignale bzw. die Störsprecher aus dem Zielsignal komplett herauszufiltern. An dieser Stelle ist eine Verbesserung der Trennungsergebnisse durch die Anwendung der sogenannten Zeit-Frequenz-Maskierung möglich. Die Hauptidee dabei ist es, durch die Analyse der geschätzten Quellsignale und extrahierten Mischungsmatrizen<sup>1</sup>, die Stellen in den Signalen zu detektieren, an denen die bleibenden Übersprecher aktiv sind und diese dann durch eine geeignete Maske zu minimieren.

In der vorliegenden Arbeit werden zwei Klassen von Verfahren zur Lösung des Permutationsproblems vorgestellt. Ein weiterer Forschungsaspekt liegt auf der Entwicklung neuer Algorithmen zur Zeit-Frequenz-Maskierung. Unter Anderem liegt der Schwerpunkt auf der Adaption der einkanaligen Störgeräuschreduktionsverfahren (z.B. Ephraim-Malah-Filter usw.) zu den mehrkanaligen Situationen.

Die Arbeit ist folgendermaßen gegliedert: im zweiten Kapitel werden die mathematischen Grundlagen erläutert, die für die vorliegende Arbeit sowie für das Verständnis von der Problematik der blinden Quellentrennung und der einkanaligen Geräuschreduktion von Bedeutung sind. Vor allem wird hier ein Einblick in die Informationstheorie sowie die Grundlagen der Statistik gegeben.

In Kapitel drei werden die wichtigsten Aspekte der blinden Quellentrennung betrachtet und es wird ein Überblick über die bekanntesten Algorithmen der ICA gegeben, die später in der vorliegenden Arbeit für die Trennung der Audiosignale benutzt werden.

In Kapitel vier wird zuerst die blinde Quellentrennung der Audiosignale im Frequenzbereich diskutiert und anschließend ein Überblick über die Methoden der

---

<sup>1</sup>Dabei werden Richtcharakteristika, Amplitudendifferenz, mehrkanalige Voice-Activity-Detection sowie psychoakustische Merkmale der Signale für die Analyse benutzt.

---

Permutationskorrektur gegeben. Darauf folgend werden zwei Klassen eigener Algorithmen entwickelt und die dazu notwendigen theoretischen Grundlagen erläutert.

Kapitel fünf gibt einen Überblick über die bekannten Maskierungsverfahren. Die drei bekanntesten Verfahren werden dann genauer vorgestellt. Anschließend wird in diesem Kapitel ein Maskierungsverfahren basierend auf der Annahme der zeitlichen Dominanz einer extrahierten Quelle pro Zeit-Frequenz-Schritt entwickelt. Weiterhin wird ein Überblick über die Methoden der einkanaligen Spektralen Subtraktion gegeben und über die Möglichkeiten der Kombination dieser Verfahren mit der mehrkanaligen Zeit-Frequenz-Maskierung bzw. über die möglichen Erweiterungen dieser Verfahren auf eine mehrkanalige Situation diskutiert. Zum Schluss wird ein Algorithmus zur mehrkanaligen Spektralen Subtraktion unter Berücksichtigung der Sprachwahrscheinlichkeiten definiert.

Kapitel sechs gibt einen Überblick über die Bewertungsmethoden für die blinde Quellentrennung von Audiosignale. Diese werden dann im Kapitel sieben bei der Auswertung der Ergebnisse eingesetzt. Die Auswertung erfolgt anhand realer Raumaufnahmen. Außerdem werden in diesem Kapitel die bekannten Algorithmen zur Permutationskorrektur sowie zur Zeit-Frequenz-Maskierung mit den in dieser Arbeit entwickelten Algorithmen verglichen.

Kapitel acht fasst die gewonnenen Erkenntnisse zusammen.



## 2. Mathematische Grundlagen

Die in der vorliegenden Arbeit betrachteten Methoden der mehrkanaligen Signalverarbeitung basieren auf den Konzepten der Informationstheorie und Statistik. Diese werden nachfolgend kurz vorgestellt, da sie für das Verständnis der Thematik essenziell wichtig sind.

### 2.1. Grundlagen der Statistik

#### 2.1.1. Zufallsvariablen

Um die grundlegenden Begriffe der Statistik einführen zu können, benötigt man zuerst ein paar Definitionen.

**Definition 2.1.1.** *Eine Menge aller möglichen Ergebnisse eines Zufallsexperiments heißt ein Wahrscheinlichkeitsraum und wird mit  $\Omega$  bezeichnet.*

Die Wahrscheinlichkeit  $P$  eines Ereignisses  $A \in \Omega$  wird über seine relative Häufigkeit definiert [BSMM00]. Wird jedem zufälligen Ereignis  $A$  eines Zufallsexperiments  $\Omega$  eine reelle Zahl zugeordnet, so spricht man von einer Zufallsvariablen  $X$ :

$$X : \Omega \rightarrow \mathbb{R}. \quad (2.1)$$

Das Verhalten einer Zufallsvariablen wird durch ihre Verteilungsfunktion beschrieben. Diese gibt an, mit welcher Wahrscheinlichkeit die Zufallsvariable  $X$  einen Wert kleiner oder gleich einem konkreten Wert  $x$  annimmt:

$$F_X(x) = P(X \leq x), \quad -\infty < x < \infty. \quad (2.2)$$

Besteht der Wertebereich der Zufallsvariablen aus endlich oder abzählbar unendlich vielen Werten, so spricht man von einer *diskreten Zufallsgröße*. Stammen die Werte aus einem oder mehreren Intervallen von  $\mathbb{R}$ , so spricht man von einer *kontinuierlichen Zufallsgröße*. Im Folgenden werden die kontinuierlichen Zufallsgrößen betrachtet. Für eine kontinuierliche Zufallsvariable lässt sich die *Wahrscheinlichkeitsdichte*

bzw. *Verteilungsdichtefunktion* (engl. *Probability Density Function, PDF*) definieren als:

$$f_X(x) = \frac{dF_X(x)}{dx}. \quad (2.3)$$

Die Wahrscheinlichkeitsdichte gibt die Wahrscheinlichkeit an, dass eine stetige Zufallsvariable  $X$  in einem Intervall  $[a, b]$  liegt.

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx. \quad (2.4)$$

Ausgehend von dieser Gleichung nimmt die Verteilungsfunktion aus der Gleichung 2.2 (auch als *kumulative* Verteilungsfunktion bekannt) für die kontinuierlichen Zufallsgrößen folgende Form an

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt. \quad (2.5)$$

### 2.1.2. Momente

Eine Verteilungsfunktion wird vor allem durch ihre statistischen Momente beschrieben. Diese werden nachfolgend definiert.

**Definition 2.1.2.** Sei  $X$  eine Zufallsvariable mit der Verteilungsdichtefunktion  $f_X(x)$ . Dann berechnet sich der Erwartungswert einer eindeutigen Funktion  $g(X)$  als

$$E[g(X)] = \int_X g(x) f_X(x) dx. \quad (2.6)$$

Für den Fall  $g(X) = X$  erhält man der Erwartungswert von  $X$ .

$$E[X] = \mu_X = \int_{-\infty}^{\infty} x f_X(x) dx. \quad (2.7)$$

**Definition 2.1.3.** Sei  $X$  eine Zufallsvariable und  $n \in \mathbb{N}$  eine natürliche Zahl. Dann bezeichnet man

$$\mathbf{m}_{n,X} = E[X^n] \quad (2.8)$$

als *Moment  $n$ -ter Ordnung* von  $X$ , und

$$\mu_{n,X} = E[(X - E[X])^n] = E[(X - \mu_X)^n] \quad (2.9)$$

als *zentrales Moment  $n$ -ter Ordnung* von  $X$ .

Momente werden oft mit Hilfe der *momentengenerierenden Funktion*  $\Phi_X(\omega)$  angegeben [NP83]. Diese wird durch konjugiert komplexe Fouriertransformierte der Verteilungsdichtefunktion gebildet

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} f_X(x) dx = E[e^{i\omega x}]. \quad (2.10)$$

Dann ergibt sich ein Moment  $n$ -ter Ordnung als

$$\mathbf{m}_n = (-i)^n \left. \frac{d^n \Phi_X(\omega)}{d\omega^n} \right|_{\omega=0}. \quad (2.11)$$

Die momentengenerierende Funktion wird oft als eine alternative Darstellungsweise einer Verteilung benutzt, besonders in den Fällen, an dem die Verteilungsdichtefunktion eine zu komplexe Form aufweist.

Eine besonders große Rolle spielen in der Statistik die ersten vier zentralen Momente (*Erwartungswert, Varianz, Schiefe* und *Wölbung/Kurtosis*).

### 2.1.3. Kumulanten

Eine weitere Gruppe der Kenngrößen einer Verteilung bilden die *Kumulanten* (auch *Semiinvarianten*). Im Gegensatz zu Momenten können die Kumulanten nicht direkt aus der Verteilungsdichtefunktion bestimmt werden und werden über eine *kumulatengenerierende Funktion*  $\Psi_X(\omega)$  definiert [NP83]

$$\Psi_X(\omega) = \ln [\Phi_X(\omega)]. \quad (2.12)$$

Daraus berechnet sich die Kumulante  $n$ -ter Ordnung als

$$\text{Cum}[x^n] = (-i)^n \left. \frac{d^n \Psi_X(\omega)}{d\omega^n} \right|_{\omega=0}, \quad (2.13)$$

bzw. im multivariaten Fall

$$\text{Cum}[x_1^{n_1}, \dots, x_r^{n_r}] = (-i)^n \left. \frac{\partial^n \Psi_X(\omega_1, \dots, \omega_r)}{\partial \omega_1^{n_1} \dots \partial \omega_r^{n_r}} \right|_{\omega_1=\dots=\omega_r=0}. \quad (2.14)$$

### 2.1.4. Verteilungsdichten

Im Folgenden werden ein paar für diese Arbeit wichtigen Verteilungen betrachtet.

### 2.1.4.1. Gaußverteilung

Die wahrscheinlich am weitesten verbreitete ist die Gauß- bzw. die Normalverteilung. Die Verteilungsdichtefunktion der Normalverteilung einer stetigen Zufallsvariablen  $X$  ist gegeben durch

$$f_X(x|\mu_X, \sigma_X) = \mathcal{N}(X|\mu_X, \sigma_X) \quad (2.15)$$

$$= \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{-\frac{(x - \mu_X)^2}{2\sigma_X^2}\right\}, \quad (2.16)$$

wobei  $\mu_X$  für den Erwartungswert und  $\sigma_X^2$  für die Varianz stehen. Diese Form sieht im multivariaten Fall folgendermaßen aus [CT91]:

$$f_{\mathbf{X}}(\mathbf{x}|\mu_{\mathbf{X}}, \Sigma) = \mathcal{N}(\mathbf{X}|\mu_{\mathbf{X}}, \Sigma) \quad (2.17)$$

$$= \frac{1}{(2\pi)^{N/2} \det\{\Sigma\}^{-1/2}} \times \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_{\mathbf{X}})^T \Sigma^{-1}(\mathbf{x} - \mu_{\mathbf{X}})\right\}. \quad (2.18)$$

Dabei ist  $N$  die Dimension der Verteilung,  $\mu_{\mathbf{X}}$  der vektorielle Mittelwert und  $\Sigma$  die Kovarianzmatrix von  $\mathbf{x}$

$$\Sigma = E[(\mathbf{x} - \mu_{\mathbf{X}})(\mathbf{x} - \mu_{\mathbf{X}})^T]. \quad (2.19)$$

### 2.1.4.2. Generalisierte Gaußverteilung

Die oben angegebene Gaußverteilung lässt sich auf eine breitere Palette der Verteilungen erweitern [VA89]. Die Gleichung 2.16 verändert sich dann zu

$$f_X(x|\mu_X, \sigma_X, \beta_X) = \frac{\beta_X}{2a\Gamma(1/\beta_X)} \exp\left\{-\left|\frac{x - \mu_X}{a}\right|^{\beta_X}\right\}. \quad (2.20)$$

Der Skalierungsparameter  $a$  wird definiert als

$$a = \sigma_X \sqrt{\frac{\Gamma(1/\beta_X)}{\Gamma(3/\beta_X)}}, \quad (2.21)$$

und die Gammafunktion ist gegeben durch

$$\Gamma(z) = \int_0^\infty u^{z-1} e^{-u} du. \quad (2.22)$$

Der  $\beta_X$ -Parameter beschreibt dabei die Form der Verteilung und kann z.B.

durch die Anwendung der Momentenmethode, die im Lösen der Gleichung

$$\left( \frac{\mu_X^2}{\sigma_X^2} \right) = \frac{\Gamma^2(2/\beta_X)}{\Gamma(1/\beta_X)\Gamma(3/\beta_X)}, \quad \frac{\mathbf{m}_{1,X}^2}{\mathbf{m}_{2,X}} = \frac{E^2[X]}{E[X^2]} = \frac{\Gamma^2(2/\beta_X)}{\Gamma(1/\beta_X)\Gamma(3/\beta_X)}, \quad (2.23)$$

resultiert [DMGFRD03, Pra05], ermittelt werden. Das Problem bei der Bestimmung des Parameters  $\beta_X$  mit der Momentenmethode ist, dass die Gleichung 2.23 keine geschlossene Lösung hat und dementsprechend eine Näherung braucht. In der Literatur wurden mehrere Lösungsvorschläge gemacht. In [KN04] wurden dazu Look-up Tabellen benutzt und in [DMGFRD03] wurde eine Approximation für die inverse Funktion von 2.23 angegeben. Alternativ kann der Parameter  $\beta_X$  durch die Maximum-Likelihood-Schätzung bestimmt werden [BF05, LL00].

Für einen Vektor  $\mathbf{X} = (X_1, X_2, \dots, X_N)$  wird die multivariate generalisierte Gaußverteilung definiert als

$$f_{\mathbf{X}}(\mathbf{x} | \mu_{\mathbf{X}}, \Sigma, \beta_{\tilde{\mathbf{X}}}) = \det\{\Sigma\}^{-1/2} \prod_{i=1}^N \left[ \frac{\beta_{\tilde{X}_i}}{(2a_{\tilde{X}_i} \Gamma(1/\beta_{\tilde{X}_i}))} \exp \left\{ - \left| \frac{\tilde{x}_i}{a_{\tilde{X}_i}} \right|^{\beta_{\tilde{X}_i}} \right\} \right] \quad (2.24)$$

mit  $\Sigma$  als eine symmetrische positiv definite Matrix (z.B. Kovarianzmatrix von  $\mathbf{x}$ ),

$$\tilde{\mathbf{x}} = \Sigma^{-1/2}(\mathbf{x} - \mu), \quad (2.25)$$

und  $a_{\tilde{X}_i}$  und  $\beta_{\tilde{X}_i}$  als Verteilungsparameter für  $\tilde{X}_i$  [BF05].

Die Generalisierte Gaußverteilung stellt eine Verallgemeinerung für viele symmetrische Verteilungsfunktionen dar. Mit dieser lassen sich z.B. die Laplace-Verteilung<sup>1</sup> ( $\beta = 1$ ), die Gaußverteilung ( $\beta = 2$ ) und die Gleichverteilung<sup>2</sup> ( $\beta \rightarrow \infty$ ) darstellen. Die Abbildung 2.1 zeigt die unterschiedlichen Verläufe der Generalisierten Gaußverteilung in Abhängigkeit vom Parameter  $\beta$ .

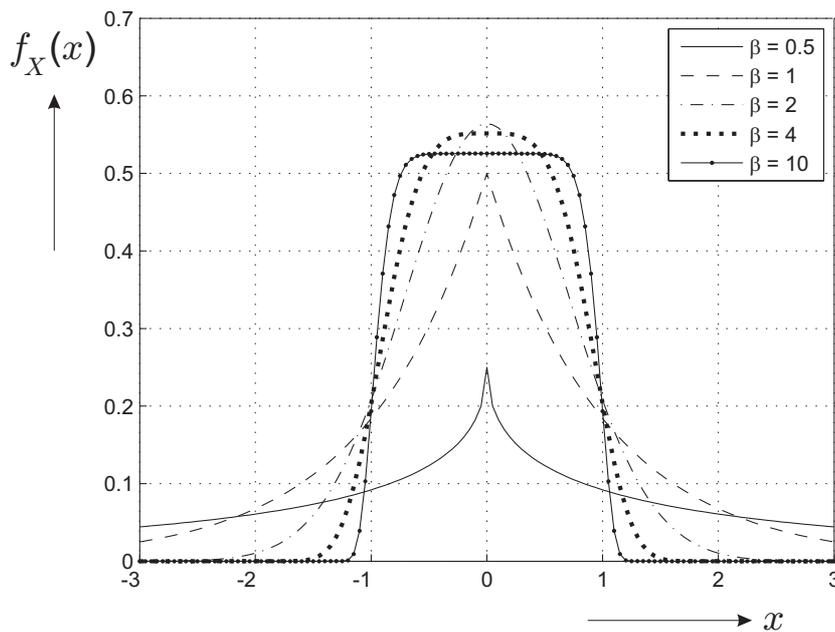
<sup>1</sup>Eine Zufallsvariable  $X$  ist Laplace-verteilt mit dem Mittelwert  $\mu_X$  und dem Varianz  $\sigma_X > 0$ , wenn sie die Wahrscheinlichkeitsdichte

$$f_X(x | \mu_X, \sigma_X) = \frac{1}{2\sigma_X} \exp \left\{ - \frac{|x - \mu_X|}{\sigma_X} \right\} \quad (2.26)$$

besitzt.

<sup>2</sup>Eine Zufallsvariable  $X$  bezeichnet man als gleichverteilt auf dem Intervall  $[a, b]$ , wenn Wahrscheinlichkeitsdichte  $f_X(x)$  gegeben ist als

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{sonst} \end{cases} . \quad (2.27)$$



**Abbildung 2.1.:** Abhängigkeit der Generalisierten Gaußverteilung vom Parameter  $\beta$  (für alle Kurven wurde  $a = 1$  gewählt).

### 2.1.4.3. Exponentialverteilung

Eine stetige Zufallsgröße  $X$  unterliegt der Exponentialverteilung mit Parameter  $\lambda$ , wenn sie die Wahrscheinlichkeitsdichte

$$f_X(x|\lambda) = \begin{cases} 0 & x < 0 \\ \lambda \exp\{-\lambda x\} & \text{sonst} \end{cases} \quad (2.28)$$

besitzt. Das Parameter  $\lambda$  wird geschätzt durch

$$\lambda = \frac{N}{\sum_{i=1}^N x_i}. \quad (2.29)$$

### 2.1.4.4. Rayleigh-Verteilung

Die Verteilungsdichtefunktion der Rayleigh-Verteilung für eine stetige Zufallsvariable  $X$  ist gegeben durch

$$f_X(x|\sigma_X) = \begin{cases} 0 & x < 0 \\ \frac{x}{\sigma_X^2} \exp\left\{-\frac{x^2}{2\sigma_X^2}\right\} & \text{sonst} \end{cases}, \quad (2.30)$$

wobei  $\sigma_X$  die Standardabweichung von  $X$  ist, geschätzt durch

$$\sigma_X = \sqrt{\frac{1}{2N} \sum_{i=1}^N x_i^2}. \quad (2.31)$$

Ähnlich wie im Falle der Normalverteilung existieren Möglichkeiten der Verallgemeinerung der Rayleigh-Verteilung. Zwei solche Verteilungsfunktionen, die für die vorliegende Arbeit relevant sind, werden im Folgenden diskutiert.

#### 2.1.4.5. Weibull-Verteilung

Die Verteilungsdichtefunktion der Weibull-Verteilung für eine Zufallsvariable  $X$  ist gegeben durch

$$f_X(x | \lambda, k) = \begin{cases} 0 & x < 0 \\ \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} \exp\left\{-\left(\frac{x}{\lambda}\right)^k\right\} & \text{sonst} \end{cases}, \quad (2.32)$$

mit  $\lambda > 0$  und  $k > 0$  als Parameter, die die Form der Weibull-Verteilung festlegen. Die Bestimmung der beiden Parameter erfolgt z.B. durch die Maximum-Likelihood-Schätzung [Coh65].

Durch die Weibull-Verteilung werden z.B. die Exponentialverteilung (für  $k = 1$ ) und die Rayleigh-Verteilung (für  $k = 2$ ) vereint. Die Abbildung 2.2 zeigt die Dichtefunktionen der Weibull-Verteilung für verschiedene Werte von  $k$ .

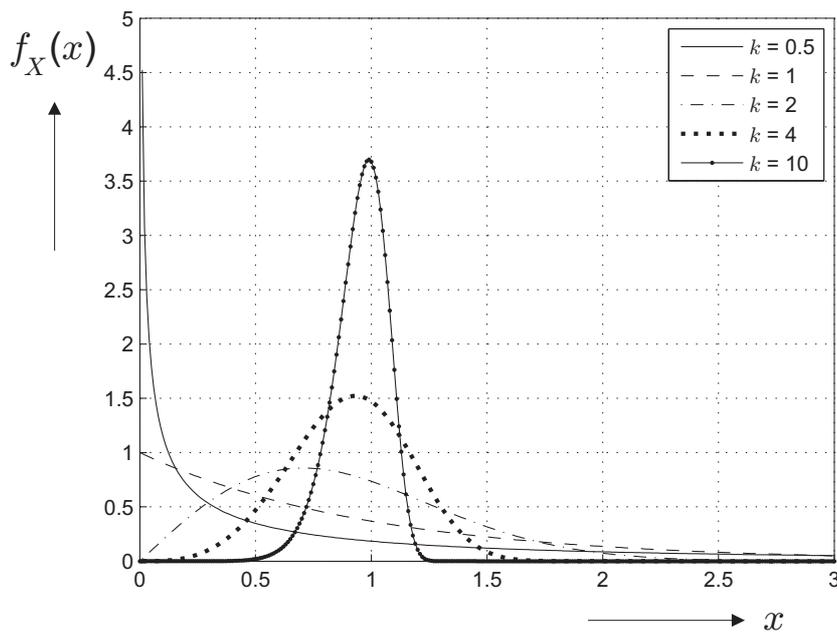
Die korrespondierende multivariate  $N$ -dimensionale Weibull-Verteilung wird definiert als [DV00, ZN05]

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^N \frac{\beta + i - 1}{\lambda_i} k_i \left(\frac{x_i}{\lambda_i}\right)^{k_i - 1} \exp\left\{-\left(\frac{x_i}{\lambda_i}\right)^{k_i}\right\} \times \left[ \sum_{j=1}^N \exp\left\{-\left(\frac{x_j}{\lambda_j}\right)^{k_j}\right\} - N + 1 \right]^{-(\beta + N)} \quad (2.33)$$

mit  $x_i > 0$  und  $\beta > 0$ .

#### 2.1.4.6. Rice-Verteilung

Eine weitere Verallgemeinerung der Rayleigh-Verteilung, die für die späteren Betrachtungen wichtig ist, ist die Rice-Verteilung [Ric45]. Die Verteilungsdichtefunktion der Rice-Verteilung für eine Zufallsvariable  $X$  ist gegeben durch



**Abbildung 2.2.:** Abhängigkeit der Weibull-Verteilung vom Parameter  $k$  (für alle Kurven wurde  $\lambda = 1$  gewählt).

$$f_X(x | \nu, \sigma) = \frac{x}{\sigma^2} \exp \left\{ -\frac{x^2 + \nu^2}{2\sigma^2} \right\} I_0 \left( \frac{x\nu}{\sigma^2} \right), \quad (2.34)$$

mit  $\nu > 0$  und  $\sigma > 0$  als Parameter, die die Form der Rice-Verteilung festlegen und  $I_0(\cdot)$  als modifizierten Bessel-Funktion erster Ordnung. Wie im Falle der Weibull-Verteilung, können auch die Parameter der Rice-Verteilung durch die Maximum-Likelihood-Schätzung bestimmt werden [SDSVD98].

Für  $\nu = 0$  ergibt sich aus der Gleichung (2.34) die Rayleigh-Verteilung.

Die Abbildung 2.3 zeigt die Dichtefunktionen der Rice-Verteilung für verschiedene Werte von  $\nu$  bei einem festen  $\sigma$  ( $\sigma = 1$ ).

## 2.2. Grundlagen der Informationstheorie

Die Informationstheorie ist eine mathematische Theorie, die auf der Wahrscheinlichkeitstheorie und Statistik basiert. Die Grundlagen der Informationstheorie wurden in den 40-er Jahren vom Claude Elwood Shannon geliefert.

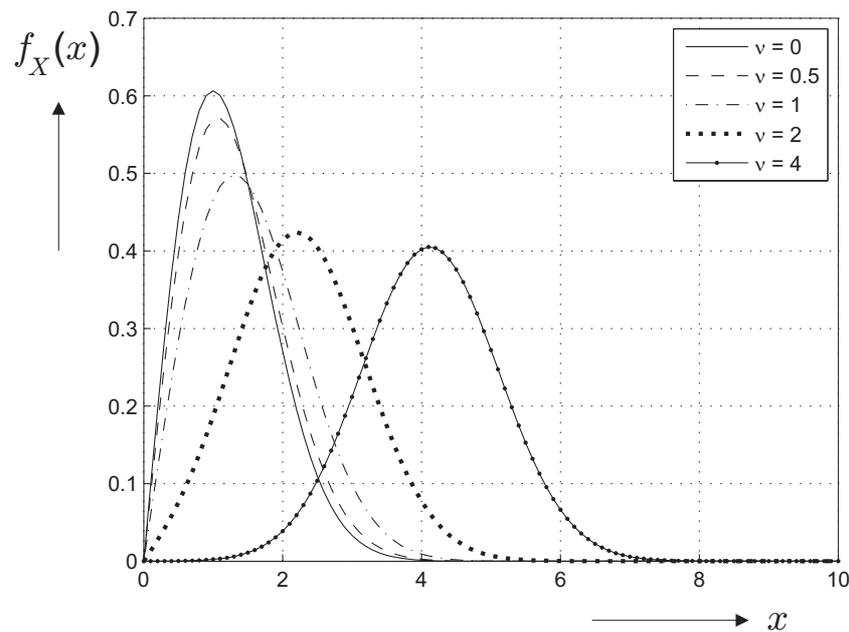


Abbildung 2.3.: Abhängigkeit der Rice-Verteilung vom Parameter  $\nu$  (für alle Kurven wurde  $\sigma = 1$  gewählt).

### 2.2.1. Entropie

Der Begriff Entropie stammt aus der Thermodynamik und wurde von Shannon und Weaver in der Informationstheorie aufgenommen und gilt seit dem als ein Maß für den Informationsgehalt einer Datenquelle.

#### 2.2.1.1. Entropiedefinition nach Shannon

Die Definition von Shannon basiert auf der Überlegung, dass je geringer die Auftretenswahrscheinlichkeit eines Ereignisses ist, desto mehr Information damit verbunden ist. Für eine diskrete Zufallsvariable  $X$  wird die Shannonsche Entropie definiert als

$$H(X) = - \sum_{\forall i} P(x_i) \log P(x_i). \quad (2.35)$$

Im Falle der kontinuierlichen Zufallsvariablen wird die *differentielle Entropie* eingeführt.

**Definition 2.2.1.** Sei  $X$  eine Zufallsvariable mit der Verteilungsdichtefunktion  $f_X(x)$  und sei  $g : \mathcal{X} \rightarrow \mathbb{R}$  eine eindeutige Funktion mit  $g(x) = -\log f_X(x)$ . Die

differentielle Entropie  $H(X)$  des Zufallsvektors  $X$  ist definiert durch

$$H(X) = E[g(X)] = - \int_{-\infty}^{\infty} f_X(x) \log f_X(x) dx. \quad (2.36)$$

falls  $E[g(X)]$  existiert.

### 2.2.1.2. Entropiedefinition nach Rényi

Eine verallgemeinerte Definition von Entropie wurde von Alfred Rényi gegeben und ist als die Rényi Entropie bekannt [R76]. Die differentielle Entropie von der Ordnung  $\alpha$ , mit  $\alpha \geq 0$ , einer Zufallsvariablen  $X$  mit einer Verteilungsdichte  $f_X(x)$  auf dem Definitionsbereich  $\mathbb{X}$  ist nach Rényi gegeben durch

$$H_\alpha^R(X) = \frac{1}{1-\alpha} \log \left( \int_{\mathbb{X}} f_X^\alpha(x) dx \right) \quad (2.37)$$

$$= \frac{\alpha}{1-\alpha} \log \left( \sqrt[\alpha]{\int_{\mathbb{X}} f_X^\alpha(x) dx} \right) \quad (2.38)$$

$$= \frac{\alpha}{1-\alpha} \log (\|f_X^\alpha(x)\|_\alpha). \quad (2.39)$$

Für den Fall  $\alpha \rightarrow 1$ ,  $H_\alpha^R(f(x))$  konvergiert diese Form gegen die Shannonsche Entropie [CT91, PXF00]

$$H_{\alpha \rightarrow 1}^R(X) = \lim_{\alpha \rightarrow 1} \frac{1}{1-\alpha} \log \left( \int_{\mathbb{X}} f_X^\alpha(x) dx \right) \quad (2.40)$$

Setzt man an dieser Stelle die Regel von L'Hospital ein und betrachtet, dass für den Ausdruck unter dem Integral  $\frac{d}{dt} a^t = a^t \log a$  gilt, so erhält man

$$H_{\alpha \rightarrow 1}^R(X) = \lim_{\alpha \rightarrow 1} \frac{\frac{\partial}{\partial \alpha} \log \left( \int_{\mathbb{X}} f_X^\alpha(x) dx \right)}{\frac{\partial}{\partial \alpha} (1-\alpha)} \quad (2.41)$$

$$= - \lim_{\alpha \rightarrow 1} \frac{\int_{\mathbb{X}} f_X^\alpha(x) \log f_X(x) dx}{\int_{\mathbb{X}} f_X^\alpha(x) dx}. \quad (2.42)$$

Da für  $\alpha = 1$   $\int_{\mathbb{X}} f_X^\alpha(x) dx = 1$  gilt, wird die Gleichung (2.42) zu

$$H_{\alpha \rightarrow 1}^R(X) = - \lim_{\alpha \rightarrow 1} \int_{\mathbb{X}} f_X^\alpha(x) \log f_X(x) dx \quad (2.43)$$

$$= - \int_{\mathbb{X}} f_X(x) \log f_X(x) dx = H(X). \quad (2.44)$$

Ähnlich der Entropie einer Zufallsvariablen  $X$  wird die Verbundentropie eines Vektors  $\mathbf{X}$  definiert als

$$H_\alpha^R(\mathbf{X}) = \frac{1}{1 - \alpha} \log \left( \int_{\mathbb{X}} f_{\mathbf{X}}^\alpha(\mathbf{x}) d\mathbf{x} \right), \quad (2.45)$$

mit  $f_{\mathbf{X}}(\mathbf{x})$  als einer multivariaten Verteilungsdichtefunktion von  $\mathbf{X}$ .

### 2.2.1.3. Entropiedefinition nach Tsallis

Sowohl die Shannonsche als auch die Rényi-Entropie sind extensiv. Das heißt, wenn zwei Variablen  $X$  und  $Y$  statistisch unabhängig sind, so gilt für ihre Verbundentropie

$$H_\alpha^R(X, Y) = H_\alpha^R(X) + H_\alpha^R(Y). \quad (2.46)$$

Die Idee der Extensivität wurde in der klassischen Informationstheorie aus der Thermodynamik adoptiert („Bei den unabhängigen Systemen werden die Entropien addiert“) und wird meistens als erwünscht angesehen. Im Gegensatz dazu zeigen die modernen Entwicklungen die Tendenz, eher auf die nicht extensive (oder auch anti-extensive) Entropiedefinitionen zurück zu greifen [HC67, Fur06, MFA<sup>+</sup>08]. Als solche wird meistens die Tsallis-Entropie verwendet. Die Tsallis-Entropie der Ordnung  $q$ , mit  $q \in \mathbb{R}, q \geq 0$ , einer Zufallsvariablen  $X$  mit einer Verteilungsdichte  $f_X(x)$  auf dem Definitionsbereich  $\mathbb{X}$  ist gegeben durch

$$H_q^T(X) = \frac{1}{q - 1} \left( 1 - \int_{\mathbb{X}} f_X^q(x) dx \right) \quad (2.47)$$

$$= \frac{1}{q - 1} \left( 1 - \|f_X^q(x)\|_q \right). \quad (2.48)$$

Für die Definition der Tsallis-Entropie wird öfters ein so genannter  $q$ -Logarithmus eingeführt [Yam02, MFA<sup>+</sup>08]

$$\ln_q x = \begin{cases} \ln x, & q = 1 \\ \frac{1-x^{1-q}}{q-1}, & \text{sonst} \end{cases}. \quad (2.49)$$

Mit

$$\ln_q 1/x = \frac{1 - (\frac{1}{x})^{1-q}}{q-1} = \frac{1 - x^{q-1}}{q-1} = -x^{q-1} \ln_q x \quad (2.50)$$

wird die Tsallis-Entropie zu

$$H_q^T(X) = \frac{1}{q-1} \left( 1 - \int_{\mathbb{X}} f_X^q(x) dx \right) \quad (2.51)$$

$$= \frac{1}{q-1} \left( \int_{\mathbb{X}} f_X(x) dx - \int_{\mathbb{X}} f_X^q(x) dx \right) \quad (2.52)$$

$$= \frac{1}{q-1} \left( \int_{\mathbb{X}} f_X(x) (1 - f_X^{q-1}(x)) dx \right) \quad (2.53)$$

$$= \int_{\mathbb{X}} f_X(x) \ln_q \left( \frac{1}{f_X(x)} \right) dx \quad (2.54)$$

$$= - \int_{\mathbb{X}} f_X(x) f_X(x)^{q-1} \ln_q f_X(x) dx \quad (2.55)$$

$$= - \int_{\mathbb{X}} f_X(x)^q \ln_q f_X(x) dx. \quad (2.56)$$

Setzt man  $q = 1$ , so wird der obere Ausdruck zu

$$H_1^T(X) = - \int_{\mathbb{X}} f_X(x) \ln_1 f_X(x) dx \quad (2.57)$$

$$= - \int_{\mathbb{X}} f_X(x) \ln f_X(x) dx \quad (2.58)$$

$$= H(X). \quad (2.59)$$

Somit stellt die Tsallis-Entropie neben der Entropiedefinition nach Rényi eine weitere Generalisierung der Shannon-Entropie dar.

Betrachtet man nun die Eigenschaften des  $q$ -Logarithmus (so wie sie in [Yam02] angegeben sind)

$$\ln_q \prod_{k=1}^n x_k = \sum_{k=1}^n (1-q)^{k-1} \sum_{i_k > \dots > i_1=1}^n \ln_q x_{i_1} \dots \ln_q x_{i_k}, \quad (2.60)$$

so wird die Verbundentropie zweier unabhängiger Zufallsvariablen  $X$  und  $Y$

zu

$$H_q^T(X, Y) = H_q^T(X) + H_q^T(Y) + (1 - q)H_q^T(X)H_q^T(Y). \quad (2.61)$$

Zwischen Tsallis- und Rényi-Entropien besteht folgender Zusammenhang

$$H_q^R(X) = \frac{1}{1 - q} \log \{1 + (1 - q)H_q^T(X)\}. \quad (2.62)$$

### 2.2.2. Generalisierte Rényi-Divergenz

Ein Kriterium für die Ähnlichkeit zweier Verteilungen wird in der Informationstheorie durch die *Generalisierte Rényi-Divergenz*, auch bekannt als  $\alpha$ -*Divergenz*, definiert. Typischerweise wird dieses Maß benutzt, um eine Distanz zwischen einer wahren Verteilungsdichte und einer Hypothese darzustellen. Die Generalisierte Rényi-Divergenz der Ordnung  $\alpha$ , mit  $\alpha \geq 0$ , zwischen zwei Verteilungsdichten  $f_X(x)$  und  $g_X(x)$  ist gegeben durch [R76]

$$D_\alpha^R(f_X(x)||g_X(x)) = \frac{1}{\alpha - 1} \log \left( \int f_X^\alpha(x) g_X^{1-\alpha}(x) dx \right). \quad (2.63)$$

Spezialfälle der Gleichung (2.63) [HMMG01] sind

- Bhattacharyya-Koeffizient

$$D_{1/2}^R(f_X(x)||g_X(x)) = -2 \log \left( \int \sqrt{f_X(x)g_X(x)} dx \right), \quad (2.64)$$

- Kullback-Leibler-Divergenz [HMMG01]

$$D_1^R(f_X(x)||g_X(x)) = \int f_X(x) \log \frac{f_X(x)}{g_X(x)} dx, \quad (2.65)$$

- Log-Distanz<sup>3</sup>

$$D_2^R(f_X(x)||g_X(x)) = \log E \left[ \frac{f_X(x)}{g_X(x)} \right], \quad (2.68)$$

---

3

$$D_2^R(f_X(x)||g_X(x)) = \log \left( \int f_X^2(x) g_X^{-1}(x) dx \right) = \log \left( \int f_X(x) \frac{f_X(x)}{g_X(x)} dx \right), \quad (2.66)$$

mit  $h_X(x) = \frac{f_X(x)}{g_X(x)}$

$$D_2^R(f_X(x)||g_X(x)) = \log \left( \int f_X(x) h_X(x) dx \right) = \log E [h_X(x)] = \log E \left[ \frac{f_X(x)}{g_X(x)} \right]. \quad (2.67)$$

- und Log-Maximum-Ratio-Distanz

$$D_{\infty}^R(f_X(x)||g_X(x)) = \log \sup_x \frac{f_X(x)}{g_X(x)}. \quad (2.69)$$

Wie oben schon erwähnt ist, beschreibt die Generalisierte Rényi-Divergenz die Ähnlichkeit zwischen zwei Verteilungen. Je kleinere Werte die Rényi-Divergenz annimmt, desto ähnlicher sind die Verteilungen. Vorgreifend kann man sagen, dass der Hauptvorteil bei der Verwendung von der Generalisierten Rényi-Divergenz in ihrer geringen Rechenlast liegt. Andererseits muss man betonen, dass dieses Maß asymmetrisch

$$D_{\alpha}^R(f_X(x)||g_X(x)) \neq D_{\alpha}^R(g_X(x)||f_X(x)). \quad (2.70)$$

und nicht begrenzt ist, so dass bei den Berechnungen unendliche Werte entstehen können.

Wird an der Stelle ein symmetrisches Kriterium gewünscht, so ist dies möglich durch

$$D_{\alpha, \text{Symm}}^R(f_X(x)||g_X(x)) = D_{\alpha}^R(f_X(x)||g_X(x)) + D_{\alpha}^R(g_X(x)||f_X(x)), \quad (2.71)$$

oder alternativ durch die so genannte  $\lambda$ -Divergenz ( $0 \leq \lambda \leq 1$ )

$$\begin{aligned} D_{\lambda, \alpha}^R(f_X(x)||g_X(x)) &= \lambda D_{\alpha}^R(f_X(x)||\lambda f_X(x) + (1 - \lambda)g_X(x)) \\ &\quad + (1 - \lambda) D_{\alpha}^R(g_X(x)||\lambda g_X(x) + (1 - \lambda)f_X(x)). \end{aligned} \quad (2.72)$$

### 2.2.3. Generalisierte Tsallis-Divergenz

Ähnlich der Generalisierten Rényi-Divergenz wird die Generalisierte Tsallis-Divergenz zwischen zwei Verteilungen definiert. Die Generalisierte Tsallis-Divergenz der Ordnung  $q$ , mit  $q \geq 0$ , zwischen zwei Verteilungsdichten  $f_X(x)$  und  $g_X(x)$  ist gegeben durch [Tsa88]

$$D_q^T(f_X(x)||g_X(x)) = \frac{1}{1 - q} \left( 1 - \int f_X^q(x) g_X^{1-q}(x) dx \right) \quad (2.73)$$

$$= \frac{1}{1 - q} \int f_X(x) \left( 1 - \left( \frac{f_X(x)}{g_X(x)} \right)^{q-1} \right) dx \quad (2.74)$$

$$= -\frac{1}{q - 1} \int f_X(x) \left( 1 - \left( \frac{g_X(x)}{f_X(x)} \right)^{1-q} \right) dx \quad (2.75)$$

$$= -\int f_X(x) \ln_q \frac{g_X(x)}{f_X(x)} dx \quad (2.76)$$

Wie im Falle der Rényi-Divergenz, wird auch die Tsallis-Divergenz bei  $q = 1$  zur Kullback-Leibler-Divergenz.

$$D_1^T(f_X(x)||g_X(x)) = - \int f_X(x) \ln_1 \frac{g_X(x)}{f_X(x)} dx \quad (2.77)$$

$$= \int f_X(x) \log \frac{f_X(x)}{g_X(x)} dx. \quad (2.78)$$

### 2.2.4. Gemeinsame Information

Ein Sonderfall der Kullback-Leibler-Divergenz bildet die *gemeinsame Information* (engl. *mutual information*). Dieses Kriterium wird für einen Vektor  $\mathbf{X} = [X_1, \dots, X_n]$  der Zufallsvariablen aufgestellt, um auf die statistischen Abhängigkeiten zwischen den Variablen zu schließen

$$I(\mathbf{X}) = D_1 \left( f_{\mathbf{X}}(\mathbf{x}) \left\| \left\| \prod_{i=1}^n f_{X_i}(\mathbf{x}) \right. \right. \right) \quad (2.79)$$

$$= \int f_{\mathbf{X}}(\mathbf{x}) \log \frac{f_{\mathbf{X}}(\mathbf{x})}{\prod_{i=1}^n f_{X_i}(x_i)} d\mathbf{x} \quad (2.80)$$

$$= \int f_{\mathbf{X}}(\mathbf{x}) \log f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} - \int f_{\mathbf{X}}(\mathbf{x}) \log \prod_{i=1}^n f_{X_i}(x_i) d\mathbf{x} \quad (2.81)$$

$$= \sum_{i=1}^n H_1(X_i) - H_1(\mathbf{X}). \quad (2.82)$$

Die gemeinsame Information ist also eine Differenz zwischen der Summe der Einzelentropien der Zufallsvariablen und der Verbundentropie.

### 2.2.5. Jensen-Rényi-Divergenz

Wählt man in der Gleichung (2.72)  $\lambda = 0,5$  und  $\alpha = 1$ , so erhält man als Spezialfall der  $\lambda$ -Divergenz die *Jensen-Shannon-Divergenz* (auch bekannt als *Informationsradius*)

$$D^{JS}(f_X(x)||g_X(x)) = \frac{1}{2}D_1(f_X(x)||h(x)) + \frac{1}{2}D_1(g_X(x)||h(x)). \quad (2.83)$$

mit  $h(x) = \frac{1}{2}(f_X(x) + g_X(x))$  und  $D_1$  als Kullback-Leibler-Divergenz aus der Gleichung (2.65). Setzt man den Ausdruck für die Kullback-Leibler-Divergenz in die

Gleichung (2.83) ein, so wird diese zu:

$$\begin{aligned}
 D^{JS}(f_X(x)||g_X(x)) &= \frac{1}{2} \int f_X(x) \log \frac{f_X(x)}{h(x)} dx + \frac{1}{2} \int g_X(x) \log \frac{g_X(x)}{h(x)} dx \\
 &= \frac{1}{2} \int f_X(x) \log f_X(x) dx + \frac{1}{2} \int g_X(x) \log g_X(x) dx \\
 &\quad - \frac{1}{2} \int f_X(x) \log h(x) dx - \frac{1}{2} \int g_X(x) \log h(x) dx \\
 &= -\frac{1}{2} H_1(X|f_X(x)) - \frac{1}{2} H_1(X|g_X(x)) \\
 &\quad - \frac{1}{2} \int [f_X(x) + g_X(x)] \log h(x) dx \\
 &= H_1(X|h(x)) - \frac{1}{2} H_1(X|f_X(x)) - \frac{1}{2} H_1(X|g_X(x)) \\
 &= H_1\left(X \left| \frac{f_X(x) + g_X(x)}{2} \right.\right) \\
 &\quad - \frac{H_1(X|f_X(x)) + H_1(X|g_X(x))}{2}, \tag{2.84}
 \end{aligned}$$

mit  $H_1(X|f_X(x))$  als Shannonsche Entropie der Zufallsvariablen  $X$ , wenn für  $X$  eine Verteilungsdichte  $f_X(x)$  angenommen wird. Diese Form kann auch für mehrere Verteilungen verallgemeinert werden:

$$\begin{aligned}
 D^{JS}(f_{X,1}(x)||f_{X,2}(x)||\dots||f_{X,n}(x)) \\
 = H_1\left(X \left| \frac{1}{n} \sum_{i=1}^n f_{X,i}(x) \right.\right) - \frac{1}{n} \sum_{i=1}^n H_1(X|f_{X,i}(x)). \tag{2.85}
 \end{aligned}$$

In [BHHK01] wurde die Form (2.85) für  $H_\alpha^R(X)$  erweitert, so dass ein Maß Namens *Jensen-Rényi-Divergenz* entstand:

$$\begin{aligned}
 D_\alpha^{JR}(f_{X,1}(x)||f_{X,2}(x)||\dots||f_{X,n}(x)) \\
 = H_\alpha^R\left(X \left| \sum_{i=1}^n \pi_i f_{X,i}(x) \right.\right) - \sum_{i=1}^n \pi_i H_\alpha^R(X|f_{X,i}(x)), \tag{2.86}
 \end{aligned}$$

wobei  $\pi_i$  einen Gewichtungsfaktor darstellt<sup>4</sup>.

Die Jensen-Rényi-Divergenz beruht auf der Generalisierten Rényi-Divergenz und

---

<sup>4</sup>Es sei anzumerken, dass für die Definition der Jensen-Rényi-Divergenz von Ben Hamza et al. die Gleichung (2.83) nicht stimmt, d.h.

$$D_\alpha^{JR}(f_X(x)||g_X(x)) \neq \frac{1}{2} D_\alpha(f_X(x)||h(x)) + \frac{1}{2} D_\alpha(g_X(x)||h(x)), \tag{2.87}$$

da die Gleichung (2.86) aus der Form (2.85) abgeleitet wurde. Jedoch hat diese die Form der Jensen-Rényi-Divergenz, wie es in [MFA<sup>+</sup>08] gezeigt wurde, und hat in der Signalverarbeitung ihren Einsatz gefunden [BHHK01, KKEP07].

kann als ihre Erweiterung angesehen werden. Im Gegensatz zur Generalisierten Rényi-Divergenz aber ist die Jensen-Rényi-Divergenz symmetrisch und immer begrenzt. Andererseits, auf Grund der Tatsache, dass Verteilungen indirekt über ihren Mittelwert  $\frac{1}{n} \sum_{i=1}^n f_{X,i}(x)$  verglichen werden, kann die Jensen-Rényi-Divergenz als eine Alternative zur gemeinsamen Information angesehen werden [MFA<sup>+</sup>08]. Tatsächlich, wie in [HHK03] gezeigt wurde, weisen beide Maße ähnliche Eigenschaften auf.

### 2.2.6. Jensen-Tsallis-q-Divergenz

Eine weitere Verallgemeinerung von Jensen-Shannon-Divergenz wurde in [BR82] unter Verwendung einer konkaven Funktion  $\phi(x)$

$$D^J(f_{X,1}(x)||f_{X,2}(x)||\dots||f_{X,n}(x)) = \phi\left(\frac{1}{n} \sum_{i=1}^n f_{X,i}(x)\right) - \sum_{i=1}^n \pi_i \phi(f_{X,i}(x)). \quad (2.88)$$

gegeben. Die Funktion  $\phi(x)$  wird dabei als  $\phi$ -Entropie bezeichnet. Setzt man voraus, dass die Funktion  $\phi(x)$  q-konkav ist <sup>5</sup>, so lässt sich eine alternative Variante der Jensen-Divergenz unter Verwendung der Tsallis-Entropie definieren, nämlich die *Jensen-Tsallis-q-Divergenz* [MFA<sup>+</sup>08]:

$$\begin{aligned} D_\alpha^{JT}(f_{X,1}(x)||f_{X,2}(x)||\dots||f_{X,n}(x)) \\ = H_q^T\left(X \left| \sum_{i=1}^n \pi_i f_{X,i}(x)\right.\right) - \sum_{i=1}^n \pi_i^q H_q^T(X|f_{X,i}(x)). \end{aligned} \quad (2.90)$$

### 2.2.7. Negentropie

Eine weitere wichtige Größe in der Informationstheorie ist die *Negentropie*. Die Negentropie ist ein Maß für die Abweichung einer Verteilung mit gegebener Varianz und gegebenem Mittelwert von der Gaußverteilung derselben Parameter.

$$H_{Neg,\alpha}(X) = H_\alpha^R(X|\mathcal{N}(x|\mu_X, \sigma_X)) - H_\alpha^R(X|f_X(x|\mu_X, \sigma_X)). \quad (2.91)$$

$\mathcal{N}(X|\mu_X, \sigma_X)$  steht dabei für die Normalverteilungsdichte mit dem Mittelwert  $\mu_X$  und der Standardabweichung  $\sigma_X$ .

<sup>5</sup>Eine Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  ist q-konkav, wenn für beliebige  $x, y \in \mathbb{R}$  und  $\lambda \in [0,1]$  gilt

$$f(\lambda x + (1 - \lambda)y) \geq \lambda^q f(x) + (1 - \lambda)^q f(y). \quad (2.89)$$

## 2.3. Kernfunktionen und ihre Anwendung

Im Bereich des Maschinellen Lernens, besonders mit der Wiederentdeckung der Support-Vector-Maschinen und der Entwicklung der Kernel-PCA, wurden in den letzten Jahren zahlreiche Verfahren entwickelt, die mittels einer nichtlinearen Abbildung  $\Phi$  die Berechnungen in einem hochdimensionalen Merkmalsraum durchführen. Der entscheidende Vorteil bei der Verwendung der Kernfunktionen ist die Möglichkeit, die linearen Berechnungsmethoden bzw. die linearen Modelle mit einer breiten Palette der nichtlinearen Entscheidungsfunktionen zu verbinden. Im Folgenden wird die Theorie der Kernfunktionen kurz vorgestellt.

Die Grundidee der Kern-basierten Verfahren besteht darin, die zu bearbeitenden Daten erst in einen Merkmalsraum  $\mathcal{E}$  durch eine nichtlineare Abbildung  $\Phi : \mathbb{R}^N \rightarrow \mathcal{E}$  zu transformieren und dann lineare Algorithmen im Merkmalsraum anzuwenden. Dieser Vorgang ist durch die Hoffnung motiviert, lineare Eigenschaften der Daten im Merkmalsraum zu gewinnen (wie z.B. lineare Trennbarkeit).

### 2.3.1. Kernel-Trick

Formuliert man nun die Algorithmen im Merkmalsraum so, dass die Abbildung  $\Phi(\mathbf{x})$  nur in Form von Skalarprodukten auftritt und diese durch eine Funktion  $\kappa : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$

$$\kappa(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle \quad (2.92)$$

effektiv bestimmt werden können, so muss die Abbildung  $\Phi(\mathbf{x})$  nicht unbedingt bekannt sein. Diese Vorgehensweise ist als *Kernel-Trick* bekannt und wurde von [ABR64] vorgestellt. Die Funktion  $\kappa$  heißt dann die *Kernfunktion* oder *Kernel*.

Die Definition von Kernfunktionen wurde im Theorem von Mercer gegeben [Mer09, SS01].

**Theorem 2.3.1.** Mercer Theorem: *Für eine kontinuierliche Funktion  $K(\mathbf{x}, \mathbf{y})$  existiert eine Entwicklung der Art*

$$K(\mathbf{x}, \mathbf{y}) = \sum_{h=0}^{\infty} \lambda_h \phi_h^T(\mathbf{x}) \phi_h(\mathbf{y}) \quad (2.93)$$

mit  $\lambda_h > 0$  dann und nur dann, wenn  $K(\mathbf{x}, \mathbf{y})$  symmetrisch (d.h.  $K(\mathbf{x}, \mathbf{y}) = K(\mathbf{y}, \mathbf{x})$ )

und nicht negativ (semi)definit ist, d.h.

$$\int_{\mathbf{Y}} \int_{\mathbf{X}} K(\mathbf{x}, \mathbf{y}) g(\mathbf{x}) g(\mathbf{y}) d\mathbf{x} d\mathbf{y} \geq 0 \quad (2.94)$$

für jede Funktion  $g : \mathcal{X} \rightarrow \mathbb{R}$ , für die gilt

$$\int_{\mathbf{X}} g^2(\mathbf{x}) d\mathbf{x} < \infty. \quad (2.95)$$

Aus praktischen Gründen werden für die Kernfunktionen öfters folgende Bedingungen gestellt [Bau01]:

•

$$\int_{-\infty}^{\infty} \kappa(\mathbf{x}) x d\mathbf{x} = 0 \quad (2.96)$$

• und

$$\int_{-\infty}^{\infty} \kappa(\mathbf{x}) d\mathbf{x} = 1. \quad (2.97)$$

D.h. man wählt zentrierte und auf 1 normierte Kerne aus.

Einige wichtige Kernfunktionen sind in der Tabelle 2.1 aufgeführt.

### 2.3.2. Korrentropie

Ausgehend vom Kernel-Trick wurde in [SPP06] ein Abstandsmaß zwischen zwei mehrdimensionalen Zufallsvariablen  $\mathbf{X}$  und  $\mathbf{Y}$

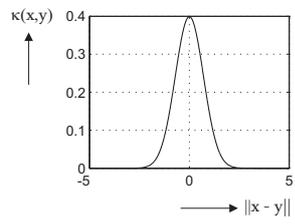
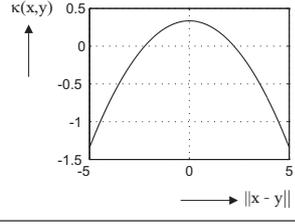
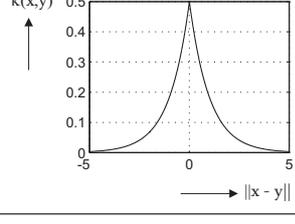
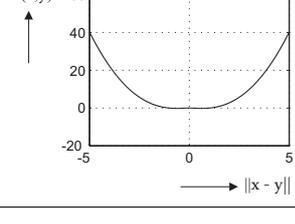
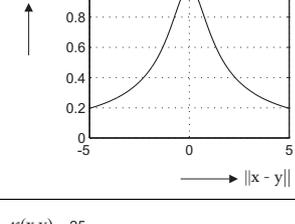
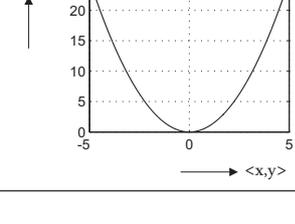
$$V(\mathbf{X}, \mathbf{Y}) = E_{\mathbf{X}\mathbf{Y}}[\kappa(\mathbf{X}, \mathbf{Y})] \quad (2.98)$$

$$= \int_{\mathbf{Y}} \int_{\mathbf{X}} \kappa(\mathbf{x}, \mathbf{y}) f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \quad (2.99)$$

definiert. Dieser wird als *Korrentropie* (engl. *Correntropy*) bezeichnet. Wie in [SPP06, LPP07] gezeigt wurde, bildet die Korrentropie eine Verallgemeinerung der Korrelation und wird deswegen in der Literatur unter anderem auch als GCF (engl. *Generalized Correlation Function*) bezeichnet.

Die Korrentropie besitzt unter anderem zwei wichtige Eigenschaften:

**Tabelle 2.1.:** Einige Kernfunktionen im Überblick [Sil86, SS01].

Kern	$\kappa(\mathbf{x}, \mathbf{y})$	Graph
Gauß-Kern	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\ \mathbf{x}-\mathbf{y}\ ^2}{2\sigma^2}\right\}$	
Epanechnikov-Kern	$\begin{cases} \frac{3}{4} \frac{(1-\frac{1}{5}\ \mathbf{x}-\mathbf{y}\ ^2)}{\sqrt{5}}, & \ \mathbf{x}-\mathbf{y}\  < \sqrt{5} \\ 0, & \text{sonst} \end{cases}$	
Laplace-Kern	$\frac{1}{2} \exp\{-\ \mathbf{x}-\mathbf{y}\ \}$	
Thin Plane Spline Kern	$\ \mathbf{x}-\mathbf{y}\ ^{2n} \log \ \mathbf{x}-\mathbf{y}\ , n > 0$	
Inv. multiquadr. Kern	$\frac{1}{\sqrt{\ \mathbf{x}-\mathbf{y}\ ^2+c^2}}, c > 0$	
Polynomialer Kern	$\langle \mathbf{x}, \mathbf{y} \rangle^d, d > 0$	

- die Korrentropie ist symmetrisch, d.h.  $V(\mathbf{X}, \mathbf{Y}) = V(\mathbf{Y}, \mathbf{X})$ ,
- die Korrentropie ist positiv und beschränkt (z.B. für Gaußkern  $0 < V(\mathbf{X}, \mathbf{Y}) \leq 1/\sqrt{2\pi}\sigma$ ) und erreicht ihren Maximalwert, wenn  $\mathbf{X} = \mathbf{Y}$ .

Diese machen die Korrentropie für die spätere Betrachtungen wichtig.



# 3. Methoden der blinden Quellentrennung

Wie bereits oben erwähnt wurde, ist die blinde Quellentrennung ein Verfahren zur Rekonstruktion der Quellsignale aus den beobachteten Mischungen ohne weitere Informationen über die Quellen, Quellenkonstellation oder Mischungsart. Das Problem der blinden Quellentrennung wurde in den letzten Jahren sehr intensiv erforscht und dabei sind viele unterschiedlichen Lösungsansätze entstanden [MK03, PLKP07].

Dieses Kapitel gibt einen Überblick über die Methoden der blinden Quellentrennung für das lineare, instantane Mischungsmodell.

## 3.1. Lineares Mischungsmodell

Bei einem linearen instantanen Mischungsmodell geht man davon aus, dass  $N$  zur gleichen Zeit aktive Quellsignale  $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]$  durch ein zeitinvariantes Mischungssystem linear unter Einfluss von Rauschsignalen  $\mathbf{n}(t) = [n_1(t), \dots, n_M(t)]$  überlagert werden, so dass  $M$  Mischsignale (Beobachtungen)  $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]$  entstehen. Diesen Vorgang kann man folgendermaßen darstellen:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t). \quad (3.1)$$

Das Mischungssystem  $\mathbf{A}$  entspricht einer  $(M \times N)$ -Matrix:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{pmatrix} = (a_{ij}), \quad (3.2)$$

wobei der Eintrag  $a_{ij}$  dem Gewichtungsfaktor entspricht, mit dem das Quellsignal  $j$  am Sensor  $i$  vorliegt. Abhängig von der Anzahl der aktiven Quellsignale und

der Anzahl der vorhandenen Sensoren sind folgende Mischungstypen zu unterscheiden:

- $M > N$  *überdeterminierte Mischung*. Dadurch, dass die Anzahl der Sensoren größer als die Anzahl der Quellsignale ist, entsteht ein „Überschuss“ an Information. Dementsprechend werden die Mischungen in einen Unterraum der Dimension  $N$  transformiert, um dadurch die Rauschanteile zu minimieren. Dies wird normalerweise mittels des *Pre-Whitening* gewährleistet, was später genauer erklärt wird. Ansonsten sieht der Entmischungsprozess genauso aus, wie in einem  $M = N$  Fall.
- $M = N$  *quadratische Mischung*. Der Standardfall der blinden Quellentrennung. Viele Lösungsansätze vereinfachen sich enorm, bei der Annahme von  $M = N$ .
- $M < N$  *unterdeterminierte Mischung*. Durch Mangel an Informationen, kann das Problem nur in bestimmten Spezialfällen gelöst werden. Es existieren jedoch problemspezifische Lösungen, wie z.B. DUET [YR04], die aber unter Umständen auf starken Annahmen basieren, die nur unter bestimmten Bedingungen in Kraft treten.

Der Entmischungsprozess besteht dann darin, eine Matrix  $\mathbf{W}$  zu finden, für die gilt:

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) = \mathbf{s}(t), \quad (3.3)$$

also

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) = \mathbf{W}\mathbf{A}\mathbf{s}(t) = \mathbf{I}\mathbf{s}(t) = \mathbf{s}(t) \Rightarrow \mathbf{W} = \mathbf{A}^{-1}, \quad (3.4)$$

wobei  $\mathbf{I}$  die Einheitsmatrix ist. Die meisten der existierenden Lösungsansätze basieren auf folgenden Annahmen:

- die Quellsignale sind statistisch unabhängig,
- höchstens eine Quelle ist normalverteilt,
- das Mischsystem  $\mathbf{A}$  ist zeitinvariant und invertierbar und
- die Rauschkomponente  $\mathbf{n}$ , wenn eine angenommen wird, ist normalverteilt.

Die wahrscheinlich wichtigste Annahme bei der Bestimmung der Matrix  $\mathbf{W}$  ist die Annahme der statistischen Unabhängigkeit der gemischten Quellsignale. Die Analyse der statistischen Unabhängigkeit wird in der Literatur als ICA (Independent Component Analysis) bezeichnet und bildet das Grundprinzip der meisten Methoden der blinden Quellentrennung [HKO01, CA02].

## 3.2. Statistische Unabhängigkeit

Zwei Zufallsvariablen  $X$  und  $Y$  sind *statistisch unabhängig*, wenn das Wissen über die Verteilung der einen Variablen keine Informationen über die Verteilung der anderen Variablen liefert [HKO01]. Mathematisch wird die statistische Unabhängigkeit folgendermaßen definiert:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y). \quad (3.5)$$

Oder verbal, die Variablen sind dann und nur dann statistisch unabhängig, wenn ihre Verbundverteilungsdichte gleich dem Produkt der Randverteilungsdichten ist. Diese Definition lässt sich sehr einfach auf mehrere Zufallsvariablen erweitern:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_i f_{X_i}(x_i). \quad (3.6)$$

Die Idee bei dem Entmischungsprozess wäre nun

1. eine Kontrastfunktion  $\mathcal{J}(\mathbf{W})$  (*contrast function*) zu finden, die quantitative Aussagen über die statistische Unabhängigkeit der Variablen geben könnte,
2. durch Minimieren der Kontrastfunktion  $\mathcal{J}(\mathbf{W})$  den Parametersatz  $\mathbf{W}$  zu finden, bei dem die Statistische Unabhängigkeit am größten ist. D.h.

$$\mathbf{W} \stackrel{!}{=} \arg \min_{\mathbf{W}} \mathcal{J}(\mathbf{W}). \quad (3.7)$$

Bei den meisten der unten diskutierten Methoden der ICA wird als Kontrastfunktion die gemeinsame Information aus der Gleichung 2.82 benutzt.

## 3.3. Korrelation

Ein Maß für die lineare Abhängigkeit der Zufallsvariablen bildet die Korrelation. Für zwei Zufallsvariablen  $X$  und  $Y$  ist die Korrelation definiert als:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}, \quad (3.8)$$

wobei  $\sigma_X$  und  $\sigma_Y$  die entsprechenden Standardabweichungen von  $X$  und  $Y$  sind, und  $\text{cov}(X, Y)$  die Kovarianz der beiden Variablen darstellt

$$\text{cov}(X, Y) = E[XY] - \mu_X \mu_Y. \quad (3.9)$$

Dabei bezeichnet  $E[\cdot]$  die Erwartungswertbildung, Gleichung 2.6.

Die Unkorreliertheit ist eine schwächere Bedingung als die im vorigen Abschnitt definierte statistische Unabhängigkeit und wird von der zweiten impliziert, d.h. sie bildet eine notwendige Bedingung der statistischen Unabhängigkeit. Allerdings wird Dekorrelation in vielen ICA Algorithmen als eine Vorverarbeitungsstufe benutzt, da dadurch die Problembetrachtung deutlich vereinfacht wird. Die Idee dabei ist, die gemischten Daten in einen durch die unkorrelierten Vektoren aufgespannten Raum zu transformieren, um dann nach statistisch unabhängigen Komponenten zu suchen, was einer Rotation der Daten im Unterraum entspricht. Diese Transformation erfolgt durch eine Multiplikation der korrelierten Daten mit einer *Whitening*-Matrix  $\mathbf{U}$  (in der Literatur auch öfters als *Spheringmatrix* bezeichnet). Diese berechnet sich aus der Eigenwertzerlegung der Kovarianzmatrix  $\Sigma$  der korrelierten Daten.

$$\Sigma = \mathbf{V}\Lambda\mathbf{V}^T, \quad (3.10)$$

wobei  $\mathbf{V}$  die orthogonale Matrix der Eigenvektoren und  $\Lambda$  die Diagonalmatrix der zugehörigen Eigenwerte ist. Die Whitening-Matrix  $\mathbf{U}$  ist nun gegeben als

$$\mathbf{U} = \Lambda^{-\frac{1}{2}}\mathbf{V}^T = \Sigma^{-\frac{1}{2}}. \quad (3.11)$$

Die nun durch die Whiteningmatrix  $\mathbf{U}$  transformierten Daten  $\mathbf{z}$

$$\mathbf{z} = \mathbf{U}\mathbf{x} \quad (3.12)$$

sind dekorreliert und besitzen eine einheitliche Varianz. D.h.

$$\begin{aligned} E[\mathbf{z}\mathbf{z}^T] &= E[\mathbf{U}\mathbf{x}\mathbf{x}^T\mathbf{U}^T] = \mathbf{U}E[\mathbf{x}\mathbf{x}]\mathbf{U}^T = \mathbf{U}E[\mathbf{x}\mathbf{x}]\mathbf{U}^T \\ &= \Lambda^{-\frac{1}{2}}\mathbf{V}^T E[\mathbf{x}\mathbf{x}]\mathbf{V}\Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}}\mathbf{V}^T\Sigma\mathbf{V}\Lambda^{-\frac{1}{2}} \\ &= \Lambda^{-\frac{1}{2}}\mathbf{V}^T\mathbf{V}\Lambda\mathbf{V}^T\mathbf{V}\Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}}\mathbf{I}\Lambda\mathbf{I}\Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}} = \mathbf{I}. \end{aligned} \quad (3.13)$$

Betrachtet man das Mischungsmodell aus der Gleichung 3.1, so gilt

$$\begin{aligned} \mathbf{z} = \mathbf{U}\mathbf{x} = \mathbf{U}\mathbf{A}\mathbf{s} = \mathbf{O}\mathbf{s} &\Rightarrow \\ E[\mathbf{z}\mathbf{z}^T] = \mathbf{I} = E[\mathbf{O}\mathbf{s}\mathbf{s}^T\mathbf{O}^T] &= \mathbf{O}E[\mathbf{s}\mathbf{s}^T]\mathbf{O}^T, \end{aligned} \quad (3.14)$$

mit  $\mathbf{O} = \mathbf{U}\mathbf{A}$ . Unter der Voraussetzung der statistischen Unabhängigkeit der Quellsignale und der Normierung der Varianz der Quellen auf eins gilt:

$$E[\mathbf{s}\mathbf{s}^T] = \mathbf{I} \Rightarrow \mathbf{O}E[\mathbf{s}\mathbf{s}^T]\mathbf{O}^T = \mathbf{O}\mathbf{I}\mathbf{O}^T = \mathbf{O}\mathbf{O}^T = \mathbf{I}. \quad (3.15)$$

D.h., dass die Matrix  $\mathbf{O}$  orthogonal ist [HKO01, Köh05].

## 3.4. Algorithmen der linearen instantanen ICA

Ausgehend von dem linearen Mischungsmodell aus dem Abschnitt 3.1 werden im Folgenden die grundlegenden Prinzipien der linearen ICA vorgestellt.

### 3.4.1. Infomax

Einer der ältesten Algorithmen der ICA wurde 1995 von A.J. Bell und T.J. Sejnowski vorgestellt [BS95b]. Der Algorithmus basiert auf dem so genannten *Infomax-Prinzip*, wo die gemeinsame Information zwischen dem Eingangsvektor  $\mathbf{x}$  des Modells und dem transformierten Ausgangsvektor  $\mathbf{z} = g(\mathbf{y}) = g(\mathbf{W}\mathbf{x})$  maximiert wird.

$$I(\mathbf{Z}, \mathbf{X}) = H(\mathbf{Z}) - H(\mathbf{Z}|\mathbf{X}) \quad (3.16)$$

$g(\cdot)$  bezeichnet dabei eine nichtlineare Funktion. Wie in [BS95a] gezeigt wurde, läuft die Maximierung der gemeinsamen Information auf die Maximierung der Verbundentropie  $H(\mathbf{Z})$  hinaus. Bei der Wahl der Nichtlinearität  $g(\cdot)$  so, dass  $g(\cdot)$  (bis auf einen Skalierungsfaktor) gleich der Verteilungsfunktion der entsprechenden unabhängigen Komponenten ist, wird die gemeinsame Information zwischen den Ausgängen  $\mathbf{y}$  des linearen Mischungsmodells minimiert, was der Wiederherstellung der statistischen Unabhängigkeit entspricht [NP94]. In der Realität sind jedoch die Verteilungsfunktionen der unabhängigen Komponenten unbekannt, was die richtige Wahl von  $g(\cdot)$  problematisch macht. Im worst case stimmt die berechnete Lösung nicht mit der gewünschten überein.

Die Berechnung der Matrix  $\mathbf{W}$  erfolgt adaptiv. In jedem Iterationsschritt werden die Einträge der Entmischungsmatrix in Abhängigkeit vom Gradienten der Verbundentropie  $H(\mathbf{Z})$  berechnet

$$\Delta\mathbf{W} \propto \frac{\partial H(\mathbf{Z})}{\partial \mathbf{W}} \quad (3.17)$$

mit  $H(\mathbf{Z})$  als Shannonsche Verbundentropie

$$H(\mathbf{Z}) = - \int_{-\infty}^{\infty} f_{\mathbf{z}}(\mathbf{z}) \log f_{\mathbf{z}}(\mathbf{z}) d\mathbf{z} = -E [\log(f_{\mathbf{z}}(\mathbf{z}))]. \quad (3.18)$$

Die Verbundverteilungsdichte der transformierten Variablen ergibt sich

aus

$$f_{\mathbf{z}}(\mathbf{z}) = \frac{f_{\mathbf{x}}(\mathbf{x})}{|\det\{\mathbf{J}\}|} \quad (3.19)$$

mit  $\mathbf{J}$  als Jacobi-Matrix

$$\mathbf{J} = \begin{pmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} & \dots & \frac{\partial z_1}{\partial x_N} \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} & \dots & \frac{\partial z_2}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial z_N}{\partial x_1} & \frac{\partial z_N}{\partial x_2} & \dots & \frac{\partial z_N}{\partial x_N} \end{pmatrix} = \begin{pmatrix} \frac{\partial z_i}{\partial x_j} \end{pmatrix} = \begin{pmatrix} w_{ij} \frac{\partial z_i}{\partial y_j} \end{pmatrix}. \quad (3.20)$$

Da jede Komponente  $z_i$  nur von  $y_i$  abhängt, verschwinden die Ableitungen  $\frac{\partial z_i}{\partial y_j}$  für  $i \neq j$ , also gilt

$$\mathbf{J} = \begin{pmatrix} w_{11} \frac{\partial z_1}{\partial y_1} & 0 & \dots & 0 \\ 0 & w_{22} \frac{\partial z_2}{\partial y_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{nn} \frac{\partial z_N}{\partial y_N} \end{pmatrix}. \quad (3.21)$$

Die Determinante von  $\mathbf{J}$  berechnet sich dann aus

$$\det\{\mathbf{J}\} = \det\{\mathbf{W}\} \prod_{i=1}^N \left| \frac{\partial z_i}{\partial y_i} \right|. \quad (3.22)$$

Setzt man nun die Gleichungen 3.18,3.19 und 3.22 in die Gleichung 3.17 ein, so erhält man:

$$\Delta \mathbf{W} \propto \frac{\partial H(\mathbf{Z})}{\partial \mathbf{W}} = \frac{\partial}{\partial \mathbf{W}} (-E[\log(f_{\mathbf{z}}(\mathbf{z}))]) \quad (3.23)$$

$$= \frac{\partial}{\partial \mathbf{W}} (E[\log |\det\{\mathbf{J}\}|]) - \underbrace{\frac{\partial}{\partial \mathbf{W}} (E[\log(f_{\mathbf{x}}(\mathbf{x}))])}_{=0} \quad (3.24)$$

$$= \frac{\partial}{\partial \mathbf{W}} (\log |\det\{\mathbf{W}\}|) + \frac{\partial}{\partial \mathbf{W}} \left( \log \prod_{i=1}^N \left| \frac{\partial z_i}{\partial y_i} \right| \right). \quad (3.25)$$

Der erste Term dieses Ausdruckes ergibt

$$\frac{\partial}{\partial \mathbf{W}} (\log |\det\{\mathbf{W}\}|) = \mathbf{W}^{-T}. \quad (3.26)$$

Der Zweite wird zu

$$\frac{\partial}{\partial \mathbf{W}} \left( \log \prod_{i=1}^N \left| \frac{\partial z_i}{\partial y_i} \right| \right) = \frac{\partial}{\partial \mathbf{W}} \left( \sum_{i=1}^N \log |g'_i(\mathbf{w}_i^T \mathbf{x})| \right) \quad (3.27)$$

$$= \left( \sum_{i=1}^N \frac{g''_i(y_i)}{g'_i(y_i)} x_i \right) \quad (3.28)$$

$$= -\phi(\mathbf{y}) \mathbf{x}^T \quad (3.29)$$

mit

$$\phi(\mathbf{y}) = \left[ -\frac{g''_1(y_1)}{g'_1(y_1)}, \dots, -\frac{g''_N(y_N)}{g'_N(y_N)} \right]^T. \quad (3.30)$$

Somit nimmt die Gleichung 3.25 folgende Form an:

$$\Delta \mathbf{W} \propto \mathbf{W}^{-T} - \phi(\mathbf{y}) \mathbf{x}^T. \quad (3.31)$$

In [Ama97] wurde gezeigt, dass die Konvergenzgeschwindigkeit dieses Verfahrens durch die Anwendung des natürlichen Gradienten erheblich erhöht werden kann. Die Lernregel 3.31 wird dann zu

$$\Delta \mathbf{W} \propto \frac{\partial H(\mathbf{Z})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W} = (\mathbf{W}^{-T} - \phi(\mathbf{y}) \mathbf{x}^T) \mathbf{W}^T \mathbf{W} = (\mathbf{I} - \phi(\mathbf{y}) \mathbf{y}^T) \mathbf{W}. \quad (3.32)$$

### 3.4.2. FastICA

*FastICA* ist ein Verfahren zur blinden Quellentrennung, das auf der Maximierung der Nichtgaußförmigkeit der Signale basiert. Das Verfahren wurde 1999 von A. Hyvärinen vorgestellt [Hyv99]. Die Nichtgaußförmigkeit wird an Hand von Negentropie (Gleichung 2.91) berechnet. Die Entwickler des Algorithmus gingen dabei von der folgenden Definition der gemeinsamen Information aus:

$$I(\mathbf{X}) = H_{Neg,1}(\mathbf{X}) - \sum_{i=1}^n H_{Neg,1}(X_i), \quad (3.33)$$

mit  $H_{Neg,1}$  als Negentropie aus der Gleichung 2.91 [HKO01]. D.h. die Minimierung der gemeinsamen Information am Ausgang des linearen Modells resultiert in der Suche der Richtung, in der die Nichtgaussförmigkeit der einzelnen Komponente  $X_i$  zunimmt.

Die Negentropie  $H_{Neg,1}(X_i)$  wird dabei als eine Kostenfunktion genommen und, um die Berechnungen zu vereinfachen, über die Kumulanten angenähert

[HKO01]

$$H_{Neg,1}(X) \approx \frac{1}{12}E[x^3]^2 + \frac{1}{48}\kappa_X(x)^2. \quad (3.34)$$

$\kappa_X(x)$  repräsentiert dabei die Kurtosis von  $X$

$$\kappa_X(x) = \mu_{4,X} - 3\sigma_X^4 = E[(x - E[x])^4] - 3E[(x - E[x])^2]^2. \quad (3.35)$$

Für die mittelwertfreien Signale vereinfacht sich die Gleichung zu

$$\kappa_X(x) = E[(x)^4] - 3E[(x)^2]^2. \quad (3.36)$$

Somit wird ein sequenzieller Updatealgorithmus für die Matrix  $\mathbf{W}$  (bzw. für die Vektoren der Matrix)

$$\Delta \mathbf{w}_i \propto \frac{\partial |\kappa(\mathbf{y})|}{\partial \mathbf{w}_i} \quad (3.37)$$

Wenn man dabei mittelwertfreie Signale betrachtet, wird 3.37 zu

$$\Delta \mathbf{w}_i \propto \frac{\partial |\kappa(\mathbf{w}_i^T \mathbf{x})|}{\partial \mathbf{w}_i} \quad (3.38)$$

$$= \text{sign}(\kappa(\mathbf{w}_i^T \mathbf{x})) [4E(\mathbf{x}(\mathbf{w}_i^T \mathbf{x})^3) - 12\mathbf{x}(\mathbf{x}^T \mathbf{w}_i)E((\mathbf{w}_i^T \mathbf{x})^2)] \quad (3.39)$$

$$= 4\text{sign}(\kappa(\mathbf{w}_i^T \mathbf{x})) [E(\mathbf{x}(\mathbf{w}_i^T \mathbf{x})^3) - 3\mathbf{x}\mathbf{x}^T \mathbf{w}_i E((\mathbf{w}_i^T \mathbf{x})^2)] \quad (3.40)$$

Setzt man als zusätzliche Bedingung voraus, dass die Signale  $\mathbf{x}$  dekorreliert und normiert (pre-whitened) sind und dass dasselbe für die Ausgangssignale  $\mathbf{y}$  gelten soll, so gelten an der Stelle dieselben Überlegungen, wie in der Sektion 3.3, nämlich dass  $\mathbf{x}\mathbf{x}^T = \mathbf{I}$  und  $E((\mathbf{w}_i^T \mathbf{x})^2) = \|\mathbf{w}_i\| = \mathbf{I}$  ist. Daraus folgt die folgende Abhängigkeit:

$$\Delta \mathbf{w}_i \propto \text{sign}(\kappa(\mathbf{w}_i^T \mathbf{x})) [E(\mathbf{x}(\mathbf{w}_i^T \mathbf{x})^3) - 3\mathbf{w}_i]. \quad (3.41)$$

Damit die Bedingung der Orthogonalität der Matrix  $\mathbf{W}$  erhalten bleibt, werden die neuen Vektoren  $\mathbf{w}_i$  nach jedem Updateschritt normiert:

$$\mathbf{w}_i = \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}. \quad (3.42)$$

Der Schwachpunkt des oben definierten Verfahrens ist die Näherung der Negentropie (Gl. 3.34). Das Problem liegt darin, dass die Schätzung der Statistiken höherer Ordnung störanfällig ist. Um diesem Problem vorzubeugen wird an der Stelle von

Gleichung (Gl. 3.34) eine Approximation der Negentropie über die nichtquadratische Funktionen benutzt. Unter Verwendung von  $n$  nichtquadratischen Funktionen wird die Negentropie approximiert als

$$H_{Neg,1}(X) \approx \sum_{i=1}^n k_i (E[G_i(x)] - E[G_i(\nu)])^2. \quad (3.43)$$

Hierbei ist  $k_i$  eine positive Konstante und  $\nu$  eine gaußsche Variable, die als Referenz dient. Mit  $n = 2$  und  $G_1 = y^3$  und  $G_2 = y^4$  ist diese Approximation identisch mit (3.34) und ist somit eine Generalisierung der momentbasierten Approximation.

Für den einfachsten Fall  $n = 1$  ergibt sich die Iterationsregel

$$\Delta \mathbf{w}_i \propto E[\mathbf{x}g(\mathbf{w}_i^T \mathbf{x})], \quad (3.44)$$

mit  $g(x) = G'(x)$ . Dieser Algorithmus zeichnet sich jedoch durch ein schlechteres Konvergenzverhalten aus. Deswegen werden in [HKO01] einige Verbesserungen vorgeschlagen. Man geht dabei davon aus, dass die Maxima der Negentropie mit den Optima von  $E[G(\mathbf{w}_i^T \mathbf{x})]$  zusammenfallen. Die zu optimierende Kostenfunktion wird dann zu

$$J(\mathbf{w}_i) = E[G(\mathbf{w}_i^T \mathbf{x})] - \lambda \mathbf{w}_i^T \mathbf{w}_i. \quad (3.45)$$

Dabei bezeichnet  $\lambda$  den Lagrange'schen Multiplikator. Die Optima der Funktion liegen in den Punkten, an denen der Gradient verschwindet:

$$\nabla_{\mathbf{w}_i} J(\mathbf{w}_i) = \frac{\partial}{\partial \mathbf{w}_i} (E[G(\mathbf{w}_i^T \mathbf{x})] - \lambda \mathbf{w}_i^T \mathbf{w}_i) \stackrel{!}{=} 0 \quad (3.46)$$

$$= E[\mathbf{x}g(\mathbf{w}_i^T \mathbf{x})] - \beta \mathbf{w}_i = 0 = f(\mathbf{w}_i), \quad (3.47)$$

mit  $\beta = 2\lambda = E[\mathbf{w}_i^T \mathbf{x}g(\mathbf{w}_i^T \mathbf{x})]$ . Die Gleichung 3.47 wird nun mit dem Newtonschen Näherungsverfahren gelöst<sup>1</sup>. Dazu wird eine Ableitung nach  $\mathbf{w}$  von der Funktion  $f(\mathbf{w}_i)$  genommen:

$$f'(\mathbf{w}_i) = \frac{\partial}{\partial \mathbf{w}_i} (E[\mathbf{x}g(\mathbf{w}_i^T \mathbf{x})] - \beta \mathbf{w}_i) = E[\mathbf{x}\mathbf{x}^T g'(\mathbf{w}_i^T \mathbf{x})] - \beta \mathbf{I}. \quad (3.48)$$

Unter der Bedingung der Orthogonalität der Matrix  $\mathbf{W}$  lässt sich der Term

<sup>1</sup>Das Newtonsche Näherungsverfahren ist das Standardverfahren zur numerischen Lösung von nichtlinearen Gleichungen und Gleichungssystemen [BSMM00]

$E [\mathbf{x}\mathbf{x}^T g'(\mathbf{w}_i^T \mathbf{x})]$  im letzten Ausdruck vereinfachen zu

$$E [\mathbf{x}\mathbf{x}^T g'(\mathbf{w}_i^T \mathbf{x})] \approx E [\mathbf{x}\mathbf{x}^T] E[g'(\mathbf{w}_i^T \mathbf{x})] = \mathbf{I} E [g'(\mathbf{w}_i^T \mathbf{x})]. \quad (3.49)$$

Damit ergibt sich eine diagonale Matrix für die Ableitung (3.48)

$$f'(\mathbf{w}_i) = \mathbf{I} (E [g'(\mathbf{w}_i^T \mathbf{x})] - \beta). \quad (3.50)$$

Die Iteration nach dem Newtonschen Näherungsverfahren für den Vektor  $\mathbf{w}_i$  hat dann die Form

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \frac{f(\mathbf{w}_i)}{f'(\mathbf{w}_i)} \quad (3.51)$$

$$= \mathbf{w}_i - (E [\mathbf{x}g(\mathbf{w}_i^T \mathbf{x})] - \beta \mathbf{w}_i) (\mathbf{I} (E [g'(\mathbf{w}_i^T \mathbf{x})] - \beta))^{-1} \quad (3.52)$$

$$= \mathbf{w}_i - \frac{(E [\mathbf{x}g(\mathbf{w}_i^T \mathbf{x})] - \beta \mathbf{w}_i)}{(E [g'(\mathbf{w}_i^T \mathbf{x})] - \beta)} \quad (3.53)$$

mit  $\beta = E [\mathbf{w}_i^T \mathbf{x}g(\mathbf{w}_i^T \mathbf{x})]$ . Diese Form kann durch die Multiplikation der beiden Seiten der Gleichung mit  $\beta - E [g'(\mathbf{w}_i^T \mathbf{x})]$  vereinfacht werden. Daraus ergibt sich der FastICA Algorithmus

$$\mathbf{w}^+ = E [\mathbf{x}g(\mathbf{w}_i^T \mathbf{x})] - E [g'(\mathbf{w}_i^T \mathbf{x})] \mathbf{w}_i \quad (3.54)$$

$$\mathbf{w}_{i+1} = \frac{\mathbf{w}^+}{\|\mathbf{w}^+\|}. \quad (3.55)$$

Dieser Algorithmus zeichnet sich durch schnelle Konvergenz und eine programmier-technisch einfache Realisierbarkeit aus.

### 3.4.3. Verfahren auf der Basis von Statistik höherer Ordnung

Eine weitere Möglichkeit auf Abhängigkeiten zwischen den Signalen zu schließen ist durch die Statistiken höherer Ordnung, wie Momente und Kumulanten, gegeben. Wie z.B. in [Car03] gezeigt wurde, kann die Kullback-Leibler-Divergenz zwischen zwei Verteilungsdichten  $D_1(f_{\mathbf{X}}(\mathbf{x})||g_{\mathbf{Y}}(\mathbf{y}))$  folgendermaßen durch die Kumulanten approximiert werden:

$$D_1(f_{\mathbf{X}}(\mathbf{x})||g_{\mathbf{X}}(\mathbf{x})) = \frac{1}{2} \sum_{r \geq 1} \frac{1}{r!} \sum_{i_1, \dots, i_r} (\text{Cum}[x_{i_1}, \dots, x_{i_r}] - \text{Cum}[y_{i_1}, \dots, y_{i_r}])^2. \quad (3.56)$$

Außerdem weisen die Kumulanten einige Eigenschaften auf, die sie besonders nützlich für die blinde Quellentrennung machen:

1. Für zwei statistisch unabhängige Variablenmengen  $\{X_1, X_2, \dots, X_n\}$  und  $\{Y_1, Y_2, \dots, Y_n\}$  gilt:

$$\text{Cum}[X_1 + Y_1, \dots, X_n + Y_n] = \text{Cum}[X_1, \dots, X_n] + \text{Cum}[Y_1, \dots, Y_n]. \quad (3.57)$$

2. Für eine normalverteilte Variable  $X_{\text{Gauss}}$  gilt:

$$\text{Cum}[X_{\text{Gauss}}^n] = 0, \quad (3.58)$$

für alle  $n \geq 3$ .

3. Für zwei statistisch unabhängige Variablenmengen  $\{X_1, X_2, \dots, X_n\}$  und  $\{Y_1, Y_2, \dots, Y_m\}$  gilt:

$$\text{Cum}[X_1, \dots, X_n, Y_1, \dots, Y_m] = 0. \quad (3.59)$$

Zieht man diese Eigenschaften in Betracht, so lässt sich eine Kostenfunktion für die blinde Quellentrennung erstellen, die die Kreuzkumulanten in Abhängigkeit von der Entmischungsmatrix minimiert. Basierend auf dieser Überlegung wurde in [CS93] ein Verfahren vorgestellt, bei dem die Kumulantentensoren vierter Ordnung diagonalisiert werden und somit eine Entmischungsmatrix gefunden wird. Das Verfahren ist unter dem Namen *JADE-Algorithmus*<sup>2</sup> bekannt.

Dabei wird ein Kumulantentensor eines Vektors der geweißten Daten  $\mathbf{z} = [z_1, \dots, z_N]$  (Gl. 3.12) als ein Satz Kumulantenmatrizen

$$[\mathbf{Q}(\mathbf{M})]_{ij} = \sum_{k=1}^N \sum_{l=1}^N \text{Cum}[z_i, z_j, z_k, z_l] \mathbf{M}_{kl}, \quad \text{und } 1 \leq i, j \leq N \quad (3.60)$$

betrachtet. Dabei bezeichnet  $\mathbf{M}$  eine Gewichtungsmatrix und

$$\begin{aligned} \text{Cum}[z_i, z_j, z_k, z_l] &= E[z_i, z_j, z_k, z_l] - E[z_i, z_j]E[z_k, z_l] \\ &\quad - E[z_i, z_k]E[z_j, z_l] - E[z_i, z_l]E[z_j, z_k], \end{aligned} \quad (3.61)$$

<sup>2</sup>JADE steht für *Joint Approximate Diagonalization of Eigen-matrices* (engl.)

für  $E[z_i] = 0$  für  $1 \leq i \leq N$ . Betrachtet man die Überlegung zu der Gleichung 3.14

$$\mathbf{z} = \mathbf{U}\mathbf{x} = \mathbf{U}\mathbf{A}\mathbf{s} = \mathbf{O}\mathbf{s} \quad (3.62)$$

so lässt sich  $\mathbf{Q}(\mathbf{M})$  diagonalisieren

$$\mathbf{Q}(\mathbf{M}) = \mathbf{O}\mathbf{\Lambda}\mathbf{O}^T \quad (3.63)$$

mit

$$\mathbf{\Lambda} = \begin{pmatrix} \kappa_{s_1} \mathbf{o}_1^T \mathbf{M} \mathbf{o}_1 & 0 & \cdots & 0 \\ 0 & \kappa_{s_2} \mathbf{o}_2^T \mathbf{M} \mathbf{o}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \kappa_{s_N} \mathbf{o}_N^T \mathbf{M} \mathbf{o}_N \end{pmatrix}. \quad (3.64)$$

$\kappa_{s_i}$  bezeichnet dabei die Kurtosis der Variablen  $s_i$  oder  $\kappa_{s_i} = \text{Cum}[s_i, s_i, s_i, s_i]$  und  $\mathbf{o}_i$  ist die  $i$ -te Spalte der Matrix  $\mathbf{O}$ . Somit besteht das Problem der Diagonalisierung von  $\mathbf{Q}(\mathbf{M})$  in der Bestimmung einer orthogonalen Rotationsmatrix  $\mathbf{O}$ . Wählt man  $\mathbf{M} = \mathbf{b}_l \mathbf{b}_k^T$  ( $\mathbf{b}_k$  ist ein Vektor mit einer 1 an der  $k$ -ten Stelle und Nullen an allen anderen), so ergeben sich  $N^2$  Kumulantenmatrizen  $\mathbf{Q}(\mathbf{M}_p)$ , die gleichzeitig diagonalisiert werden. Dieses erfolgt über Verbunddiagonalisierung (joint diagonalization), welche eine orthogonale Basis ermittelt, mit der alle Matrizen eines Satzes näherungsweise diagonalisiert werden können. Als Algorithmus dient z. B. eine Erweiterung des Jacobi-Algorithmus [Sch97]. Die Anwendung von Jacobi-Algorithmus ist nur möglich, weil hier von den weißen Daten ausgegangen wurde und somit über die Menge orthogonaler Matrizen optimiert wird.

### 3.4.4. Verfahren auf der Basis von Statistik zweiter Ordnung

In den bisherigen Betrachtungen hat man die statistischen Kenngrößen höherer Ordnung dazu ausgenutzt, um auf die Abhängigkeiten in den Mischungen zu schließen und basierend darauf die statistische Unabhängigkeit wiederherzustellen. Sind die Spektren der gemischten Signale ausreichend voneinander unterschiedlich, so kann die blinde Quellentrennung auch auf der Basis der Statistik zweiter Ordnung durchgeführt werden. Die diesem Prinzip zugrunde liegende Idee ist, dass die statistisch unabhängigen Signale zu keinem Zeitpunkt korrelieren und dabei aber eine spektrale Färbung aufweisen. D.h. in anderen Worten, dass die Kreuzkorrelationsmatrix der

Quellsignale zu jedem Zeitpunkt diagonal bleibt

$$\mathbf{C}(t, \tau) = E[\mathbf{s}(t)\mathbf{s}(t + \tau)^T] \quad (3.65)$$

$$= \begin{pmatrix} R_{s_1 s_1}(\tau) & 0 & \cdots & 0 \\ 0 & R_{s_2 s_2}(\tau) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{s_N s_N}(\tau) \end{pmatrix}, \quad (3.66)$$

mit  $R_{s_i s_i}(\tau)$  als Autokovarianzfunktion der Quellsignale  $s_i(t)$ . Bei der Annahme eines instantanen Mischungsmodells gilt für die Mischsignale:

$$\mathbf{C}_x(t, \tau) = E[\mathbf{x}(t)\mathbf{x}(t + \tau)^T] \quad (3.67)$$

$$= \mathbf{A} \begin{pmatrix} R_{s_1 s_1}(\tau) & 0 & \cdots & 0 \\ 0 & R_{s_2 s_2}(\tau) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{s_N s_N}(\tau) \end{pmatrix} \mathbf{A}^T. \quad (3.68)$$

Geht man von den weißen Daten aus (wie in Gleichung (3.12)), so reduziert sich das Problem auf eine Suche nach einer orthogonalen Rotationsmatrix  $\mathbf{O}$ , die Kovarianzmatrix

$$\mathbf{C}_z(t, \tau) = E[\mathbf{z}(t)\mathbf{z}(t + \tau)^T] \quad (3.69)$$

zu jedem Zeitpunkt diagonalisiert

$$\mathbf{O}^T \mathbf{C}_z(t, \tau) \mathbf{O} = \mathbf{O}^T E[\mathbf{z}(t)\mathbf{z}(t + \tau)^T] \mathbf{O} \quad (3.70)$$

$$= E[\mathbf{O}^T \mathbf{z}(t)\mathbf{z}(t + \tau)^T \mathbf{O}] \quad (3.71)$$

$$= E[\mathbf{s}(t)\mathbf{s}(t + \tau)^T]. \quad (3.72)$$

Somit ergibt sich ein Problem der simultanen Diagonalisierung mehrerer Matrizen, das schon im vorigen Abschnitt diskutiert worden ist. Wie bei JADE kann auch an dieser Stelle ein Jacobi-Verfahren für die Diagonalisierung benutzt werden. Diese Methode ist unter dem Namen *SOBI* (*Second Order Blind Identification*) bekannt [BACM97].

In [Köh99] wurde eine Verallgemeinerung dieses Konzeptes vorgeschlagen. Diese beruht auf der Verwendung der linearen Operatoren

$$\mathbf{T}(\mathbf{x}(t)) = [T_1(x_1(t)), \dots, T_N(x_N(t))]. \quad (3.73)$$

Die Anwendung dieser Operatoren auf die durch Pre-Whitening vorverarbeiteten Signale ergibt

$$\mathbf{T}(\mathbf{z}(t)) = \mathbf{T}(\mathbf{O}\mathbf{s}(t)) = \mathbf{O}\mathbf{T}(\mathbf{s}(t)). \quad (3.74)$$

Bildet man nun die Kovarianzmatrix der transformierten Signale, so erhält man

$$\mathbf{C}_z = E[\mathbf{T}(\mathbf{z}(t))\mathbf{T}(\mathbf{z}(t))^T] \quad (3.75)$$

$$= E[\mathbf{T}(\mathbf{O}\mathbf{s}(t))\mathbf{T}(\mathbf{O}\mathbf{s}(t))^T] \quad (3.76)$$

$$= \mathbf{O}E[\mathbf{T}(\mathbf{s}(t))\mathbf{T}(\mathbf{s}(t))^T]\mathbf{O}^T \quad (3.77)$$

$$= \mathbf{O}\mathbf{D}_T\mathbf{O}^T. \quad (3.78)$$

Da die linearen Operatoren die statistischen Eigenschaften der Signale nicht beeinflussen, ist die Matrix  $\mathbf{D}_T$  diagonal. Wenn an dieser Stelle mehrere verschiedene Operatoren verwendet werden, so entstehen mehrere Kovarianzmatrizen  $\mathbf{C}_i$ , die simultan zu diagonalisieren sind. Durch eine geeignete Wahl der linearen Operatoren können gewisse spektrale Eigenschaften der Signale stärker hervorgehoben werden, was die Trennungsergebnisse verbessert. In [Köh99] wurden z.B. optimale FIR-Filter und Wavelet-Transformation verwendet.

### 3.4.5. Unzulänglichkeiten der linearen blinden Quellentrennung

Allen oben beschriebenen Verfahren der blinden Quellentrennung sind zwei Einschränkungen gemeinsam.

1. Bei der Bestimmung der unabhängigen Komponenten können keine Energien der Quellsignale bestimmt werden. Für eine beliebige Diagonalmatrix  $\mathbf{D}$  lässt sich die Gleichung (3.1) zu  $\mathbf{x}(t) = \mathbf{A}\mathbf{D}\mathbf{D}^{-1}\mathbf{s}(t)$  umformen, d.h. dass die Signale  $\mathbf{D}^{-1}\mathbf{s}(t)$  ebenfalls statistisch unabhängig sind. Diese Tatsache ist auch unter der Bezeichnung *Skalierungsproblem* bekannt. Ein weit verbreiteter Weg mit dem Skalierungsproblem umzugehen, liegt in der Normierung der Varianzen der Eingangssignale auf Einheitsvarianz.
2. Die Reihenfolge der unabhängigen Komponenten kann nicht bestimmt werden. Ähnlich wie beim Skalierungsproblem zerstört eine Permutationsmatrix  $\mathbf{P}$  die Annahme der statistischen Unabhängigkeit nicht. Das wird durch die Gleichung

$$\mathbf{x}(t) = \mathbf{A}\mathbf{P}\mathbf{P}^{-1}\mathbf{s}(t) \quad (3.79)$$

ersichtlich.

Zusammenfassend kann man sagen, dass die durch die oben beschriebenen Algorithmen erhaltenen unabhängigen Signale nur bis auf eine bestimmte Skalierung und Permutation den wahren Quellsignalen gleichen

$$\mathbf{y}(t) = \mathbf{P}\mathbf{D}\mathbf{s}(t). \quad (3.80)$$

Während das Permutationsproblem bei den instantanen Mischungen keine große Rolle spielt, kann es dramatische Auswirkungen auf die Ergebnisse der Quellentrennung der konvolutiven Mischungen haben. Diese Tatsache wird im nächsten Kapitel genauer diskutiert.



## 4. Blinde Quellentrennung konvolutiver Mischungen

Im Kapitel 3 wurden das Problem der blinden Quellentrennung ausgehend vom linearen instantanen Mischungsmodell sowie die unterschiedlichen Lösungsansätze vorgestellt. In vielen realen Situationen, wie z.B. im Falle der akustischen Mischungen, ist dieses Modell jedoch nicht ausreichend, da dabei eine andere Mischungsart vorliegt. So bleiben u.a. die Signallaufzeiten zwischen den Sensoren sowie die Schallreflexionen an den Wänden unberücksichtigt. Um das Problem handhaben zu können, muß das Mischungsmodell erweitert werden, was u.U. neue Lösungsansätze benötigt. Im Falle der akustischen Mischungen liegt eine Überlagerung gefilterter (*konvolutiver*) Quellsignale vor. Bei der Entmischung der auf diese Art entstandenen Signale spielt neben dem Problem der blinden Quellentrennung an sich das oben erwähnte Permutationsproblem eine große Rolle, das einer separaten Behandlung bedarf.

Dieses Kapitel gibt einen Überblick über die Vorgehensweise bei der blinden Quellentrennung der konvolutiven Mischungen sowie über die Lösung des Permutationsproblems.

### 4.1. Konvolutives Mischungsmodell

Zuerst wird die im Kapitel 3 beschriebene Situation verallgemeinert, sodass ein neues Signalmodell erstellt wird, das von konvolutiven Mischungen ausgeht. In einem realen Raum sind  $N$  Signalquellen  $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]$  aktiv, die durch  $M$  Mikrofone aufgenommen werden, so dass ein Vektor  $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]$  entsteht. Die genaue Platzierung der Mikrofone wird an dieser Stelle zunächst nicht in Betracht gezogen. Durch die unterschiedlichen Laufzeiten, die die Quellsignale brauchen, um zu unterschiedlichen Mikrofonen zu gelangen, sowie durch die Reflexionen, die an den Wänden des Raumes entstehen, lassen sich die Mischungen  $\mathbf{x}(t)$  durch eine Faltung der Quellsignale mit unbekanntem Raumimpulsantworten  $\underline{\mathbf{a}} = (\underline{\mathbf{a}}_1 \dots \underline{\mathbf{a}}_K)$  darstellen,

mit  $\mathbf{a}_k$  als  $k$ -ter ( $k \in [1 \dots K]$ )  $M \times N$  Matrix mit Filterkoeffizienten und  $K$  als Filterlänge

$$\mathbf{x}(t) = \sum_{k=0}^{K-1} \mathbf{a}_{k+1} \mathbf{s}(t-k) + \mathbf{n}(t). \quad (4.1)$$

Der Term  $\mathbf{n}(t)$  bezeichnet dabei das additive Sensorrauschen.

Das Ziel ist nun eine Filtermatrix  $\underline{\mathbf{w}} = (\underline{\mathbf{w}}_1 \dots \underline{\mathbf{w}}_{K'})$  zu finden, so dass bei der Faltung mit den Beobachtungssignalen  $\mathbf{x}(t)$  die Quellsignale wiederhergestellt werden können

$$\mathbf{y}(t) = \sum_{k'=0}^{K'-1} \underline{\mathbf{w}}_{k'+1} \mathbf{x}(t-k'). \quad (4.2)$$

In anderen Worten muss  $\mathbf{y}(t) \approx \mathbf{s}(t)$  gelten. Die übliche Vorgehensweise bei der Trennung konvolutiver Mischungen besteht darin, die Mischsignale mit Hilfe der *Kurzzeit-Fouriertransformation*<sup>1</sup> zunächst in den Frequenzbereich zu transformieren

$$\mathbf{X}(\Omega, \tau) = \sum_{t=-\infty}^{\infty} \mathbf{x}(t) w(t-\tau R) e^{-j\Omega t}, \quad (4.3)$$

mit  $\Omega$  als Kreisfrequenz,  $\tau$  als Zeitindex im Frequenzbereich,  $w(t)$  als Fensterfunktion (z.B. Hanningfester) der Länge  $N_{FFT}$  und  $R$  als Verschiebungsfaktor, in Zeitschritten zwischen den benachbarten Fenstern [AR77]. Diese Vorgehensweise reduziert die Faltung im Zeitbereich (Gleichung 4.1) auf  $N_{FFT}/2$  komplexe Multiplikationen mit der entsprechenden Übertragungsfunktion im Frequenzbereich, so dass für jedes Frequenzband ein instantanes komplexwertiges Mischungsproblem

$$\mathbf{X}(\Omega, \tau) \approx \mathbf{A}(\Omega) \mathbf{S}(\Omega, \tau) + \mathbf{N}(\Omega, \tau) \quad (4.4)$$

entsteht. Zur Lösung können instantane ICA-Algorithmen aus dem Kapitel 3 benutzt werden, die für jedes Frequenzband eine separate Entmischungsmatrix  $\mathbf{W}(\Omega) \approx \mathbf{A}^{-1}(\Omega)$  berechnen

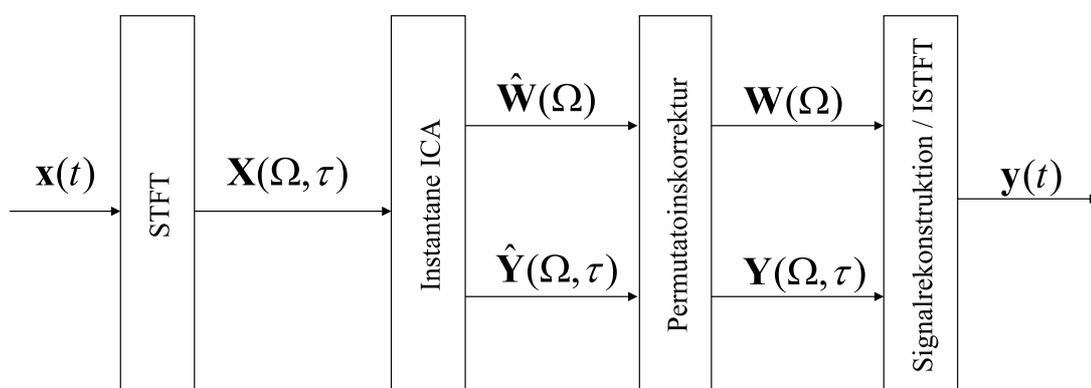
$$\hat{\mathbf{Y}}(\Omega, \tau) = \mathbf{W}(\Omega) \mathbf{X}(\Omega, \tau). \quad (4.5)$$

Da in diesem Fall die Frequenzbänder unabhängig voneinander betrachtet werden, können die berechneten Ausgangssignale  $\hat{\mathbf{Y}}(\Omega, \tau)$  beliebig untereinander vertauscht (permutiert) sein, was durch die Eigenschaften der ICA-Algorithmen zu erklären ist (siehe Abschnitt 3.4.5). Um nun konsistente Signale im Zeitbereich zu gewinnen, müssen auf diese Art entstandene Permutationen in eine konsistente Reihenfolge

---

<sup>1</sup>engl. Short Time Fourier Transformation

gebracht werden, d.h. für jede Frequenz muss eine Korrekturmatrix  $\mathbf{P}(\Omega)$  berechnet werden. Die Rekonstruktion der Signale erfolgt entweder durch die Entmischung im Frequenzbereich mit der anschließenden Rücktransformation oder durch die Rücktransformation der Matrizen  $\mathbf{W}(\Omega)$  in den Zeitbereich und der FIR-Filterung. Der Datenfluss bei diesem Ablauf ist in der Abbildung 4.1 dargestellt.



**Abbildung 4.1.:** Datenfluss bei der frequenzbasierten Entmischung konvolutiver Mischungen.

Neben der oben beschriebenen Vorgehensweise existiert eine Reihe an Verfahren, die das Problem entweder im Zeitbereich angehen [KBM<sup>+</sup>98] oder eine Implementierung im Frequenzbereich mit einer Kostenfunktion im Zeitbereich kombinieren [Lam96, LBL97].

## 4.2. Überblick über die Verfahren zur Permutationskorrektur

Wie im vorigen Abschnitt erklärt wurde, ist der Entmischungsprozess im Frequenzbereich mit einem zusätzlichen Aufwand, nämlich mit dem Permutationsproblem,

verbunden. Die Lösung des Problems wird mit einer größeren Anzahl der gemischten Signale schwieriger und unter Umständen schlechter (wenn man das Gesamtergebnis betrachtet). Der Grund dafür liegt nicht nur im faktoriellen Zuwachs der Permutationsanzahl (für  $N$  gemischte Signale kommen  $N!$  mögliche Permutationen in Frage), sondern auch darin, dass für eine größere Anzahl der Signale die ICA-Lösungen schlechter ausfallen. Im Laufe der letzten Jahre wurden für das Permutationsproblem zahlreiche Lösungswege vorgeschlagen, die hier kurz diskutiert werden. Die Lösungsansätze bei diesen Verfahren kann man grob in zwei Gruppen unterteilen:

- Lösungen basierend auf der Konsistenz der Spektren der rekonstruierten Signale.
- Lösungen basierend auf der Konsistenz der Filterkoeffizienten.

#### 4.2.1. Permutationskorrekturen basierend auf der Konsistenz der Spektren

Einer der am weitesten verbreiteten Ansätze basiert auf der Verwendung der Kreuzstatistiken. Die Permutationsmatrix wird bei diesen Verfahren aus dem folgenden Zusammenhang ermittelt:

$$\hat{\mathbf{P}}(\Omega) \stackrel{!}{=} \arg \max_{\mathbf{P}(\Omega)} \mathcal{F}(\mathbf{P}(\Omega) \hat{\mathbf{Y}}(\Omega, \tau), \hat{\mathbf{Y}}(\bar{\Omega}, \tau)). \quad (4.6)$$

Hierbei ist  $\hat{\mathbf{P}}(\Omega)$  die gesuchte Korrekturmatrix für das Frequenzband  $\Omega$ , die durch die Anwendung einer statistischen Funktion  $\mathcal{F}(\cdot, \cdot)$  bzw. Abstandsfunktion zwischen dem aktuellen Frequenzband  $\hat{\mathbf{Y}}(\Omega, \tau) = [Y_1(\Omega, \tau), \dots, Y_N(\Omega, \tau)]$  und einem Referenzband  $\hat{\mathbf{Y}}(\bar{\Omega}, \tau) = [Y_1(\bar{\Omega}, \tau), \dots, Y_N(\bar{\Omega}, \tau)]$  mit  $N$  als Anzahl der Quellsignale ermittelt wird. Die Rolle der Abstandsfunktion können z.B die Kreuzkumulanten, Kreuzkorrelation oder AM-Dekorrelation<sup>2</sup>

$$C(\hat{Y}_i(\Omega, \tau), \hat{Y}_k(\bar{\Omega}, \tau)) = E[\hat{Y}_i(\Omega, \tau), \hat{Y}_k(\bar{\Omega}, \tau)] - E[\hat{Y}_i(\Omega, \tau)]E[\hat{Y}_k(\bar{\Omega}, \tau)] \quad (4.7)$$

---

<sup>2</sup>Amplitude Modulation Decorrelation (engl.)

mit  $1 \leq i \leq N, 1 \leq k \leq N$  übernehmen [AK00, MDC00, MIZ01]. In [SAM07] wurde die Kreuzkorrelation zwischen den *Leistungsverhältnissen* (engl. *power ratio*)

$$\text{powerRatio}_i(\Omega, \tau) = \frac{\left\| \mathbf{A}_i(\Omega) \hat{Y}_i(\Omega, \tau) \right\|^2}{\sum_{k=1}^N \left\| \mathbf{A}_k(\Omega) \hat{Y}_k(\Omega, \tau) \right\|^2} \quad (4.8)$$

der entmischten Signale  $\hat{\mathbf{Y}}(\Omega, \tau)$  als ein Kriterium für die Permutaionskorrekturen benutzt.  $\mathbf{A}_i(\Omega)$  bezeichnet dabei die  $i$ -te Spalte der geschätzten Mischungsmatrix  $\mathbf{A}(\Omega)$ . Das Leistungsverhältnis bildet dabei ein Maß für die Dominanz der Quelle  $i$  in der Mischung. Der Wertebereich dieses Maßes liegt im Intervall  $[0,1]$ .  $\text{powerRatio}_i(\Omega, \tau)$  geht gegen 1, wenn der Term  $\mathbf{A}_i(\Omega)Y_i(\Omega, \tau)$  in der Dekomposition von der Gleichung 4.5 dominant ist und gegen 0, wenn ein anderes Signal  $\mathbf{A}_j(\Omega)Y_j(\Omega, \tau), (j \neq i)$  dominiert. Das Leistungsverhältnis ist besonders geeignet für die Sprachsignale, da diese von Natur aus sparse (d.h. disjunkt in der Zeit) sind.

Weitere Verfahren basieren auf der Idee, dass nahe aneinander liegende Frequenzbänder „ähnlich“ aussehen müssen, was die Verwendung von spektralen Abstandsmaßen für die Permutationskorrekturen erlaubt. Die Bestimmung der Permutationsmatrix

$$\hat{\mathbf{P}}(\Omega) \stackrel{!}{=} \arg \min_{\mathbf{P}(\Omega)} \mathcal{A}(\mathbf{P}(\Omega) \hat{\mathbf{Y}}(\Omega, \tau), \hat{\mathbf{Y}}(\bar{\Omega}, \tau)) \quad (4.9)$$

erfolgt dabei aus der Minimierung des Abstandes  $\mathcal{A}(\mathbf{P}(\Omega) \hat{\mathbf{Y}}(\Omega, \tau), \hat{\mathbf{Y}}(\bar{\Omega}, \tau))$ . Dieser wird z.B. durch den minimalen quadratischen Fehler zwischen den Einhüllenden  $\mathcal{E}(\cdot)$  der Frequenzbänder [RABH07]

$$\mathcal{A}(\mathbf{P}(\Omega) \hat{\mathbf{Y}}(\Omega, \tau), \hat{\mathbf{Y}}(\bar{\Omega}, \tau)) = \sum_N \frac{1}{T} \sum_{\tau=1}^T \left[ \mathcal{E}(\mathbf{P}(\Omega) \hat{\mathbf{Y}}(\Omega, \tau)) - \mathcal{E}(\hat{\mathbf{Y}}(\bar{\Omega}, \tau)) \right]^2 \quad (4.10)$$

oder durch die  $p$ -Norm der Distanz zwischen den Logarithmen der Beträge der Frequenzbänder [KHK04, PSB03b]

$$\begin{aligned} & \mathcal{A}(\mathbf{P}(\Omega) \hat{\mathbf{Y}}(\Omega, \tau), \hat{\mathbf{Y}}(\bar{\Omega}, \tau)) \\ &= \sum_N \left( \frac{1}{T} \sum_{\tau=1}^T \left[ \log \left| (\mathbf{P}(\Omega) \hat{\mathbf{Y}}(\Omega, \tau)) \right|^2 - \log \left| (\hat{\mathbf{Y}}(\bar{\Omega}, \tau)) \right|^2 \right]^p \right)^{1/p} \quad (4.11) \end{aligned}$$

realisiert.

In [MD03] wurde ein Verfahren präsentiert, das ausgehend von der Annahme der Laplace-Verteilung der Signale ein Maximum-Likelihood-Kriterium für die Detektion der Permutationen benutzt

$$\hat{\mathbf{P}}(\Omega) \stackrel{!}{=} \arg \min_{\mathbf{P}(\Omega)} \prod_{i,j \in \{\mathbf{P}_{ij}^k(\Omega)=1\}} \gamma_{ij}(\Omega) \quad (4.12)$$

mit  $1 \leq i \leq N, 1 \leq j \leq N$ .  $\mathbf{P}_{ij}^k(\Omega)$  bezeichnet einen Eintrag an der Stelle  $i, j$  in der  $k$ -ten Permutationsmatrix mit  $k \in [1, K]$  und  $K$  als Anzahl der möglichen Permutationen. Hierbei ist  $\gamma_{ij}(\Omega)$  ein Skalierungsparameter

$$\gamma_{ij}(\Omega) = \frac{1}{T} \sum_{\tau} \frac{|\hat{Y}_i(\Omega, \tau)|}{\beta_j(\tau)} \quad (4.13)$$

mit

$$\beta_j(\tau) = \frac{1}{N} \sum_{\Omega} |\hat{Y}_j(\Omega, \tau)|. \quad (4.14)$$

#### 4.2.2. Permutationskorrekturen basierend auf der Konsistenz der Filterkoeffizienten

Wie oben schon erwähnt wurde, kann das Permutationsproblem dadurch gelöst werden, dass man für die Konsistenz der Filterkoeffizienten sorgt. Wie in [PS00] gezeigt wurde, hängt die Forderung nach Konsistenz der Filterkoeffizienten mit der Länge des Impulsantwortes im Zeitbereich zusammen. Sprich, eine kurze Filterlänge sorgt für einen glatten Verlauf der Filterkoeffizienten im Frequenzbereich. Diese Tatsache wurde in [VL03] ausgenutzt.

Die Konsistenz der Filterkoeffizienten bedeutet, dass die Filterkoeffizienten einen glatten Verlauf aufweisen. Dieses erreicht man, wenn man die Euklidische Norm ( $\|\cdot\|$ )

$$\mathcal{D} = \sum_{i,k} \|\mathbf{A}_{ik}^P(\Omega) - \mathbf{A}_{ik}(\bar{\Omega})\| \quad (4.15)$$

zwischen den Koeffizienten der  $i$ -ten Spalte und  $k$ -ten Zeile der permutierten Matrix  $\mathbf{A}^P(\Omega) = \mathbf{P}(\Omega)\mathbf{A}(\Omega)$  im  $n$ -ten Frequenzband und den Koeffizienten der Matrix  $\mathbf{A}(\bar{\Omega})$  eines Referenzfrequenzbandes (z.B. des  $n - 1$ -ten Frequenzbandes) minimiert [BKKO01]. In [PSB03b] wird an Stelle der Euklidischen Distanz die Diagonalität

des Produktes

$$\mathbf{R}(\Omega, \bar{\Omega}) = \mathbf{A}^P(\Omega) \mathbf{A}^{-1}(\bar{\Omega}) \quad (4.16)$$

als Entscheidungskriterium für die Permutationskorrekturen benutzt.

Ausgehend von derselben Annahme, nämlich, dass die Koeffizienten der Mischungsfilter einen glatten Verlauf aufweisen, werden in [SAMM06] die amplituden- und phasennormierten Vektoren gruppiert. Die Phasennormierung der Vektoren erfolgt durch die Wahl eines Referenzsensors  $J$

$$\bar{\mathbf{A}}_{ki}(\Omega) = |\mathbf{A}_{ki}(\Omega)| \exp \left\{ j \frac{\arg(\mathbf{A}_{ki}(\Omega)/\mathbf{A}_{Ji}(\Omega))}{2\pi\Omega c^{-1}d_{\max}} \right\} \quad (4.17)$$

mit  $c$  als Schallgeschwindigkeit und  $d_{\max}$  als dem maximalen Abstand zwischen zwei Mikrofonen im Array. Anschließend wird eine Einheitsnormierung

$$\bar{\bar{\mathbf{A}}}_i(\Omega) = \bar{\mathbf{A}}_i(\Omega) / \|\bar{\mathbf{A}}_i(\Omega)\| \quad (4.18)$$

durchgeführt. Auf diese Weise erhaltene Vektoren (Basisvektoren) werden dann mit einem Clustering-Algorithmus (z.B. *k-means*) gruppiert.

Eine andere Möglichkeit die Permutationen aufzulösen, ist durch die Ordnung der Koeffizienten nach der Kosinusdistanz

$$\cos \alpha_i = \frac{(\mathbf{A}_i^P(\Omega))^H \mathbf{A}_i(\bar{\Omega})}{\|\mathbf{A}_i^P(\Omega)\| \|\mathbf{A}_i(\bar{\Omega})\|} \quad (4.19)$$

zwischen den Vektoren  $\mathbf{A}_i(\Omega)$  der geschätzten Mischungsmatrizen gegeben [AIO<sup>+</sup>03].

Als ein Kriterium für die Permutationsdetektion können unter Anderem auch räumliche Informationen ausgenutzt werden. So wird in [SMM03] ein Verfahren vorgeschlagen, das aus den geschätzten Mischungsmatrizen die Einfallswinkel der Quellsignale<sup>3</sup>

$$\theta_i(\Omega) = \arccos \frac{\arg(\mathbf{A}_{ki}(\Omega)/\mathbf{A}_{k'i}(\Omega))}{2\pi\Omega c^{-1}d_{kk'}} \quad (4.20)$$

berechnet<sup>4</sup>, nach denen die Vektoren der Mischmatrizen anschließend sortiert werden.

---

<sup>3</sup>engl. *Direction Of Arrival (DOA)*

<sup>4</sup>Die Berechnung basiert auf der Annahme des Fernfeldmodells, d.h. der Abstand zwischen der

Hierbei bezeichnet  $c$  die Schallgeschwindigkeit und  $d_{kk'}$  den Abstand zwischen den Mikrofonen  $k$  und  $k'$ .

Eine alternative Variante für die Berechnung der Einfallswinkel ist durch die Minima der Beampattern von den Entmischungsmatrizen  $\mathbf{W}(\Omega)$  gegeben. Die Beampattern lassen sich aus der Multiplikation der durch die Zeilen der Entmischungsmatrix gebildeten Filter mit einem frequenz- und richtungsabhängigen Phasenvektor  $\mathbf{r}(\Omega, \phi) = [1, \dots, r_M(\Omega, \phi)]^T$  berechnen

$$\Psi_i(\Omega, \phi) = \mathbf{W}_i(\Omega)\mathbf{r}(\Omega, \phi). \quad (4.21)$$

Im Idealfall fallen die Einfallswinkel der Quellsignale mit den Nullstellen des jeweiligen Beampatterns zusammen. Aus diesem Grund werden in [KSK<sup>+</sup>00, BKO03, MSAM03] die Nullstellen der berechneten Beampattern für die Permutatinskorrekturen gruppiert.

Eine etwas andere Vorgehensweise wurde in [Sma97] vorgestellt. Hier wird davon ausgegangen, dass die iterativen ICA-Algorithmen bei der geeigneten Initialisierung keine Permutationen produzieren. Demzufolge wird das Permutationsproblem durch eine günstige Preinitialisierung gelöst. Als Startwert für die Entmischungsmatrix  $\mathbf{W}_0(\Omega)$  im  $k$ -ten Frequenzband wird der Wert der vorher berechneten Entmischungsmatrix des  $(k - 1)$ -ten Bandes benutzt bzw. alternativ kann  $\mathbf{W}_0(\Omega)$  mit dem Wert eines Null-Beamformers initialisiert werden, der dieselben Nullrichtungen wie  $\mathbf{W}(\Omega)$  hat. Im zweiten Fall muß der Beamformer vorher ausgerechnet werden, was den Rechenaufwand nicht unbedingt reduziert.

### 4.3. Permutationskorrekturen basierend auf den informationstheoretischen Abstandsmaßen

Im diesem Abschnitt wird eine Klasse von Algorithmen zu Permutationskorrekturen vorgestellt, die auf informationstheoretischen Abstandsmaßen basiert. Wie im Abschnitt 4.2.1 erklärt wurde, können die Permutationen dadurch in eine konsistente Reihenfolge gebracht werden, dass man für die Konsistenz der Spektren der separierten Signale sorgt. Algorithmen dieser Gruppe basieren auf der Annahme, dass die benachbarten Frequenzbänder ähnliche Verläufe aufweisen und demzufolge auf der Auswertung der Ähnlichkeitsmaße zwischen den Frequenzbändern aufbauen. Als solche können auch die informationstheoretischen Abstandsmaße, wie z.B. gemeinsame

---

Signalquelle und dem Mikrofonarray  $r$  ist viel größer als der Abstand zwischen den einzelnen Mikrofonen  $d_{kk'}$ ,  $r \gg d_{kk'}$ .

Information, Generalisierte Rényi-Divergenz etc. (s. Abschnitt 2.2) benutzt werden. Jedoch benötigen diese Abstandsmaße Vorwissen über die Verteilung der Signale. Da in der vorliegenden Arbeit hauptsächlich Sprachsignale betrachtet werden, wird im Folgenden ihre Verteilung diskutiert. Die nachfolgenden Algorithmen basieren auf den in [HKKO12] veröffentlichten Betrachtungen, die während der Arbeit an den vorliegenden Dissertation entstanden sind.

### 4.3.1. Allgemeine Betrachtungen

Wie in [EM84] gezeigt wurde, können die Real- und Imaginärteile der Fouriertransformierten Sprachsignale

$$X(\Omega, \tau) = \Re\{X(\Omega, \tau)\} + j\Im\{X(\Omega, \tau)\} = X_R(\Omega, \tau) + jX_I(\Omega, \tau) \quad (4.22)$$

durch die Gaussverteilung  $\mathcal{N}$  (Abschnitt 2.1.4.1) angenähert werden. Unter Annahme der statistischen Unabhängigkeit zwischen dem Real- und Imaginärteil der Signale ergibt sich die Verbundverteilungsdichte

$$f(X_R(\Omega, \tau), X_I(\Omega, \tau)) = \mathcal{N}(\mu_{X_R}, \sigma_{X_R})\mathcal{N}(\mu_{X_I}, \sigma_{X_I}) \quad (4.23)$$

bzw.

$$\begin{aligned} f(X_R(\Omega, \tau), X_I(\Omega, \tau)) \\ = \frac{1}{2\pi\sigma_{X_R}^2\sigma_{X_I}^2} \exp\left\{-\frac{(X_R(\Omega, \tau) - \mu_{X_R})^2 + (X_I(\Omega, \tau) - \mu_{X_I})^2}{4\sigma_{X_R}^2\sigma_{X_I}^2}\right\}, \end{aligned} \quad (4.24)$$

wobei  $\mu_{X_R}$  und  $\mu_{X_I}$  die Mittelwerte und  $\sigma_{X_R}^2$  und  $\sigma_{X_I}^2$  die Varianzen der Real- und Imaginärteile des Sprachsignals bezeichnen. Nach der Transformation zu Polarkoordinaten  $(X_R, X_I)^T = (A \cos \varphi, A \sin \varphi)^T$  unter Berücksichtigung der Transformationsvorschrift

$$\begin{aligned} dX_R(\Omega, \tau)dX_I(\Omega, \tau) &= \det \frac{\partial(X_R(\Omega, \tau), X_I(\Omega, \tau))}{\partial(A(\Omega, \tau), \varphi(\Omega, \tau))} dA(\Omega, \tau)d\varphi(\Omega, \tau) \quad (4.25) \\ &= A(\Omega, \tau)dA(\Omega, \tau)d\varphi(\Omega, \tau) \quad (4.26) \end{aligned}$$

mit  $A$  als Amplitude und  $\varphi$  als Phase des Signals<sup>5</sup>, ergibt sich die Rayleighverteilung

$$f(A(\Omega, \tau), \varphi(\Omega, \tau)) = \frac{A(\Omega, \tau)}{2\pi\sigma_X^2} \exp \left\{ -\frac{A(\Omega, \tau)^2}{4\sigma_X^2} \right\} \quad (4.28)$$

für die Verteilung von Amplituden. Der Phasenverlauf  $\varphi(\Omega, \tau)$  ist unbekannt und bei der Lösung von Permutationsproblemen unbedeutend. Demzufolge wird er eliminiert, indem man die Gleichung 4.28 über die Phase mittelt:

$$\overline{f(A(\Omega, \tau))} = \int_0^{2\pi} f(A(\Omega, \tau), \varphi(\Omega, \tau)) f(\varphi(\Omega, \tau)) d\varphi(\Omega, \tau). \quad (4.29)$$

Da die wahre Verteilungsdichte von  $\varphi(\Omega, \tau)$  unbekannt ist, wird an dieser Stelle eine Gleichverteilung von  $\varphi(\Omega, \tau)$  auf dem Intervall  $[0, 2\pi]$  angenommen, d.h.  $f(\varphi(\Omega, \tau)) = 1/2\pi$ ,  $\varphi(\Omega, \tau) \in [0, 2\pi]$ . Dann wird 4.28 zu

$$\overline{f(A(\Omega, \tau))} = \frac{A(\Omega, \tau)}{\sigma_X^2} \exp \left\{ -\frac{A(\Omega, \tau)^2}{4\sigma_X^2} \right\}. \quad (4.30)$$

Die Annahme der Rayleigh-Verteilung für den Verlauf der Amplituden ist zwar sinnvoll, wenn es um Rauschreduktion im Sprachsignal geht, weil dadurch der mathematische Aufwand bei der Spektralen Subtraktion enorm reduziert wird, ist aber an sich zu ungenau. Dies kann man in der Abbildung 4.2 gut sehen. Die Ungenauigkeit der Rayleigh-Verteilung kann dann später zu schlechteren Ergebnissen bei der Permutationskorrektur führen. Somit wäre an dieser Stelle die Annahme einer genaueren Verteilungsfunktion nötig, z.B. durch eine Generalisierung der Rayleigh-Verteilung. Dies ist, wie im Abschnitt 2.1.4.5 bereits erwähnt, z.B. durch die Weibull-Verteilung

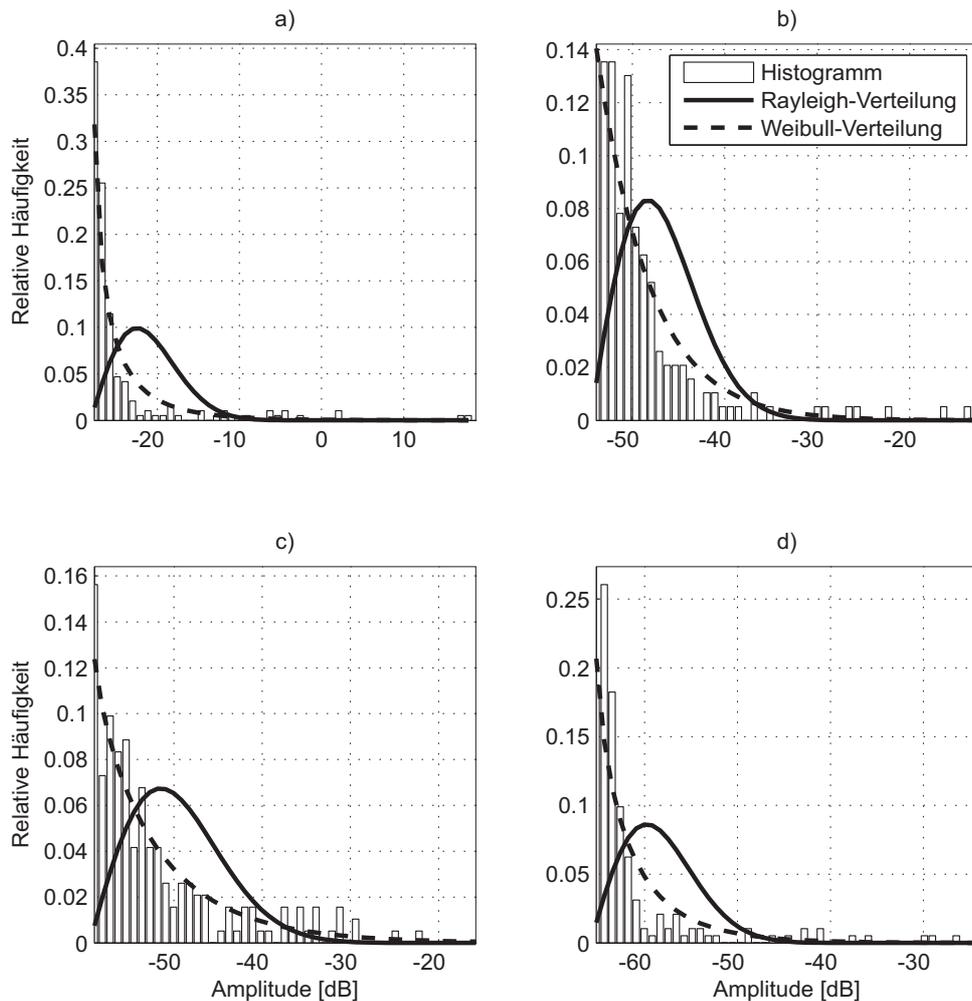
$$\overline{f(A(\Omega, \tau))} = \frac{k}{\lambda_X} \left( \frac{A(\Omega, \tau)}{\lambda_X} \right)^{k-1} \exp \left\{ -\left( \frac{A(\Omega, \tau)}{\lambda_X} \right)^k \right\} \quad (4.31)$$

mit  $\lambda_X$  und  $k$  als Weibull-Parameter aus der Gleichung (2.32) gegeben. Die Abbildung 4.2 zeigt wie die unterschiedlichen Frequenzbänder eines Sprachsignals jeweils durch Rayleigh- und Weibull-Verteilungen approximiert werden. Man kann deutlich sehen, dass die Weibull-Verteilung eine deutlich bessere Darstellung der Verteilung

---

5

$$\det \frac{\partial (X_R(\Omega, \tau), X_I(\Omega, \tau))}{\partial (A(\Omega, \tau), \varphi(\Omega, \tau))} = \begin{vmatrix} \cos \varphi(\Omega, \tau) & -A(\Omega, \tau) \sin \varphi(\Omega, \tau) \\ \sin \varphi(\Omega, \tau) & -A(\Omega, \tau) \cos \varphi(\Omega, \tau) \end{vmatrix} = A(\Omega, \tau). \quad (4.27)$$



**Abbildung 4.2.:** Dichtemodelle der Sprachsignalverteilungen unter der Annahme der Rayleigh- (schwarze Kurve) und der Weibull-Verteilungen (gestrichelte Kurve) und gemessene Histogramme der Spektralkomponenten eines Sprachsignals im Frequenzband um a) 1000 Hz, b) 2000 Hz, c) 3000 Hz und d) 4000 Hz. Das Sprachsignal wurde in einem Raum mit einer Reverberationszeit  $T_R = 159$  ms aufgenommen. Für den Vergleich wurden 10 Sekunden der Aufnahme verwendet.

der Amplituden eines Sprachsignals liefert. Somit wird für die späteren Betrachtungen die Weibull-Verteilung genommen, wenn es um die Verteilung der Amplituden bei der Permutationskorrektur geht.

Für die Beschreibung der Verteilung von Logarithmen der Amplituden der Signale kann die generalisierte Gaußverteilung (Gleichung (2.20)) benutzt werden. Diese hat

den Vorteil, dass durch sie vor allem die spitzen Verteilungsfunktionen sehr genau dargestellt werden können.

Nun wird angenommen, dass die Verteilungsdichte der Amplituden eines Signals im Frequenzband  $k$  durch die Funktion

$$f\left(\left|\hat{\mathbf{Y}}(\Omega_k, \tau)\right|\right) = \left[f\left(\left|\hat{Y}_1(\Omega_k, \tau)\right|\right), \dots, f\left(\left|\hat{Y}_N(\Omega_k, \tau)\right|\right)\right] \quad (4.32)$$

und Verteilungsdichte des  $k + 1$  Frequenzbandes durch die Funktion

$$f\left(\left|\hat{\mathbf{Y}}(\Omega_{k+1}, \tau)\right|\right) = \left[f\left(\left|\hat{Y}_1(\Omega_{k+1}, \tau)\right|\right), \dots, f\left(\left|\hat{Y}_N(\Omega_{k+1}, \tau)\right|\right)\right] \quad (4.33)$$

beschrieben werden. Um das Permutationsproblem lösen zu können, muss ein Distanzmaß  $d(\cdot, \cdot)$  definiert werden, das die Abhängigkeit zweier Verteilungsdichten misst, so dass die totale Abhängigkeit in

$$D\left(f\left(\left|\hat{\mathbf{Y}}^P(\Omega_k, \tau)\right|\right), f\left(\left|\hat{\mathbf{Y}}(\Omega_{k+1}, \tau)\right|\right)\right) = \sum_{n=1}^N d\left(f\left(\left|\hat{Y}_n^P(\Omega_k, \tau)\right|\right), f\left(\left|\hat{Y}_n(\Omega_{k+1}, \tau)\right|\right)\right), \quad (4.34)$$

resultiert. Hierbei bezeichnet

$$\hat{\mathbf{Y}}^P(\Omega_k, \tau) = \mathbf{P}(\Omega_k) \hat{\mathbf{Y}}(\Omega_k, \tau) \quad (4.35)$$

eine Permutation von  $\hat{\mathbf{Y}}(\Omega_k, \tau)$  gebildet durch die Permutationsmatrix  $\mathbf{P}(\Omega_k)$  im  $k$ -ten ( $k \in [1, N_{FFT} - 1]$ ) Frequenzband. D.h. die totale Distanz  $D$  zwischen dem permutierten Vektor  $\hat{\mathbf{Y}}^P(\Omega_k, \tau)$  und dem Referenzvektor  $\hat{\mathbf{Y}}(\Omega_{k+1}, \tau)$  wird als Summe über alle Distanzen zwischen den Paaren  $\hat{Y}_n^P(\Omega_k, \tau)$  und  $\hat{Y}_n(\Omega_{k+1}, \tau)$  berechnet.

Diese Vorgehensweise kann auch für die Verteilungen von Logarithmen der Amplituden der Signale benutzt werden. In diesem Fall werden analog die

$$f\left(\log\left|\hat{\mathbf{Y}}(\Omega_k, \tau)\right|\right) = \left[f\left(\log\left|\hat{Y}_1(\Omega_k, \tau)\right|\right), \dots, f\left(\log\left|\hat{Y}_N(\Omega_k, \tau)\right|\right)\right] \quad (4.36)$$

und

$$f\left(\log\left|\hat{\mathbf{Y}}(\Omega_{k+1}, \tau)\right|\right) = \left[f\left(\log\left|\hat{Y}_1(\Omega_{k+1}, \tau)\right|\right), \dots, f\left(\log\left|\hat{Y}_N(\Omega_{k+1}, \tau)\right|\right)\right] \quad (4.37)$$

in Betracht gezogen.

An Stelle der Ähnlichkeitsfunktion  $d(\cdot, \cdot)$  können neben den Maßen aus Abschnitten 2.2.2-2.2.7 folgende Ähnlichkeitsmaße verwendet werden:

- **Die Distanz von Chernov** [CA02] für  $0 \leq s \leq 1$  ist

$$D^C(f(x)||q(x)) = E \{f(x)^s q(x)^{1-s}\}. \quad (4.38)$$

- **Die Distanz von Kolmogorov** [CA02] für  $s > 0$  ist

$$D^K(f(x)||q(x)) = \frac{1}{2} E \{|f(x) - q(x)|^s\}. \quad (4.39)$$

- **Die generalisierte Traubert-Distanz** [CA02] für  $n \in \mathbb{N}$  ist

$$D^T(f(x)||q(x)) = E \left\{ \frac{f(x)^n + q(x)^n}{f(x)^{n-1} + q(x)^{n-1}} \right\}. \quad (4.40)$$

### 4.3.2. Definition der neuen Abstandsmaße

Die Jensen-Rényi-Divergenz  $D_{JR}(f_{X_1}(x)||g_{X_2}(x))$  aus der Gleichung (2.86) misst die Distanz zwischen den zwei Verteilungen  $f_{X_1}(x)$  und  $g_{X_2}(x)$  bezüglich eines dritten Punktes im Verteilungsraum (einer Verteilungsdichte). Als der dritte Punkt wird dabei der Mittelwert der beiden Verteilungen gewählt. Diese Vorgehensweise ist durch die Konkavität der Entropie bezüglich der Verteilungsdichten gerechtfertigt, d.h. unter anderem

$$H \left( X \left| \frac{f_X(x) + g_X(x)}{2} \right. \right) \geq \frac{H(X|f_X(x)) + H(X|g_X(x))}{2}. \quad (4.41)$$

Prinzipiell ist es möglich den Abstand bezüglich eines beliebigen Punktes anzugeben, für den man eine Aussage über die Konvexität bzw. die Konkavität der Entropien machen kann. Dies kann man erhalten, wenn man z.B. nicht über die Verteilungsfunktionen sondern über die den Verteilungsfunktionen zugrunde liegenden Variablen (in diesem Fall über die zwei Frequenzbänder) mittelt und dann die entsprechende Verteilungsfunktion als einen Bezugspunkt wählt. Betrachtet man diese Überlegung in Verbindung mit unterschiedlichen Entropiedefinitionen, wie sie im Abschnitt 2.2.1 eingeführt wurden, so führt sie zu folgender Fallunterscheidung:

- **Shannonsche Entropie.** Wie in [DCT91] gezeigt wurde, gilt für die Entropiedefinition nach Shannon folgende Ungleichung:

$$H(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda H(X) + (1-\lambda)H(Y) \quad (4.42)$$

mit  $0 \leq \lambda \leq 1$ . Der Beweis für die Ungleichung (4.42) ist in [DCT91] zu finden und basiert auf der Exponentialungleichung der Entropie<sup>6</sup>. Aus (4.42) folgt, dass die Shannonsche Entropie konkav bezüglich der gewichteten Summe der Zufallsvariablen ist. Aus dieser Tatsache lässt sich ein Abstandsmaß

$$D_\lambda(X||Y) = H(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - (\lambda H(X) + (1-\lambda)H(Y)) \quad (4.44)$$

ableiten. Wählt man  $\lambda = 0,5$ , so ergibt sich

$$D_\lambda(X||Y) = H\left(\frac{X+Y}{\sqrt{2}}\right) - \frac{H(X) + H(Y)}{2}. \quad (4.45)$$

Da für die Shannon-Entropie

$$H(aX) = H(X) + \log a \quad (4.46)$$

gilt, ist

$$H\left(\frac{X+Y}{\sqrt{2}}\right) = H(X+Y) + \log \frac{1}{\sqrt{2}} < H(X+Y). \quad (4.47)$$

Damit wird die *modifizierte Jensen-Shannon-Divergenz*

$$D^{modJS}(X||Y) = H(X+Y) - \frac{H(X) + H(Y)}{2} \quad (4.48)$$

definiert<sup>7</sup>. Für den Fall von  $n$  Verteilungen wird die obere Gleichung zu

$$D^{modJS}(X_1||X_2||\dots||X_n) = H\left(\sum_{i=1}^n X_i\right) - \frac{1}{n} \sum_{i=1}^n H(X_i). \quad (4.49)$$

- **Rényi Entropie.** Die Ungleichung (4.42) gilt für Rényi Entropie im Allgemeinen nicht. Um eine Aussage über die Konvexität der Rényi Entropie zu treffen, wird an dieser Stelle zuerst die verallgemeinerten Youngsche Ungleichung vorgestellt [Bar98].

---

6

**Theorem 4.3.1.** Exponentialungleichung der Entropie [CT91]: Für zwei unabhängige multivariate Zufallsvariablem  $\mathbf{X} = (X_1, X_2, \dots, X_N)$  und  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)$  gilt

$$2^{\frac{2}{N} H_1(\mathbf{X}+\mathbf{Y})} \geq 2^{\frac{2}{N} H_1(\mathbf{X})} + 2^{\frac{2}{N} H_1(\mathbf{Y})}. \quad (4.43)$$

Die Ungleichung (4.43) wird zur Gleichung dann und nur dann, wenn  $\mathbf{X}$  und  $\mathbf{Y}$  normalverteilt sind.

<sup>7</sup>Das Eliminieren von  $\sqrt{2}$  in  $H\left(\frac{X+Y}{\sqrt{2}}\right)$  wird an dieser Stelle vorzogen, um die neuen Abstandsmaße auf die einheitliche Form zu bringen.

**Theorem 4.3.2.** (Verallgemeinerte Youngsche Ungleichung): Seien  $f \in L^p(\mathbb{R}^N)$  und  $g \in L^q(\mathbb{R}^N)$  nicht negative Funktionen (Verteilungsdichten) und  $p, q, r > 0$ , so dass  $1/p + 1/q = 1/r + 1$ , so gilt

$$\|f * g\|_r \leq \left( \frac{C_p C_q}{C_r} \right)^N \|f\|_p \|g\|_q, \text{ wenn } p, q, r \geq 1 \quad (4.50)$$

und

$$\|f * g\|_r \geq \left( \frac{C_p C_q}{C_r} \right)^N \|f\|_p \|g\|_q, \text{ wenn } p, q, r \leq 1. \quad (4.51)$$

Hierbei ist

$$C_t = \sqrt{\frac{t^{1/t}}{\tilde{t}^{1/\tilde{t}}}}, \text{ mit } \frac{1}{t} + \frac{1}{\tilde{t}} = 1 \quad (4.52)$$

und  $*$  bezeichnet die Faltung.

Das Theorem 4.3.2 ist für die aktuelle Betrachtung in sofern wichtig, als die Verteilungsdichtefunktion der Summe zweier Zufallsvariablen  $Z = X + Y$

$$f_Z(x) = f_X(x) * f_Y(x) \quad (4.53)$$

ist [CT91, Arn04], und somit

$$H_\alpha^R(X + Y) \propto \|f_X(x) * f_Y(x)\|_\alpha \quad (4.54)$$

gilt.

Betrachtet man nun die Gleichungen (4.50) und (4.51) für den Fall, dass  $r = p$  und  $q = 1$  ist, so erhält man

$$\|f * g\|_r \leq \|f\|_r, \text{ wenn } r \geq 1 \quad (4.55)$$

und

$$\|f * g\|_r \geq \|f\|_r, \text{ wenn } r \leq 1, \quad (4.56)$$

da im Falle einer Verteilungsdichtefunktion  $\|g\|_1 = 1$  ist. Für die Rényi Entropie heißt das

$$\frac{\alpha}{1-\alpha} \log \|f_X(x) * f_Y(x)\|_\alpha \geq \frac{\alpha}{1-\alpha} \log \|f_X(x)\|_\alpha \Rightarrow \quad (4.57)$$

$$H_\alpha^R(X+Y) \geq H_\alpha^R(X). \quad (4.58)$$

Die Rényi Entropie der Summe bleibt immer größer oder gleich der Entropie der Einzelvariable, da das Vorzeichen von  $\frac{\alpha}{1-\alpha}$  für  $\alpha > 1$  kippt. Dieselbe Überlegung gilt natürlich für den Fall, dass  $r = q$  und  $p = 1$  ist, d.h.

$$H_\alpha^R(X+Y) \geq H_\alpha^R(Y). \quad (4.59)$$

Daraus folgt [BV02]

$$H_\alpha^R(X+Y) \geq \max(H_\alpha^R(X), H_\alpha^R(Y)). \quad (4.60)$$

Da  $\max(a, b) \geq (a+b)/2$  ist, gilt folgender Zusammenhang

$$H_\alpha^R(X+Y) \geq \frac{H_\alpha^R(X) + H_\alpha^R(Y)}{2}. \quad (4.61)$$

Basierend darauf wird ein Abstandsmaß als

$$D_\alpha^{modJR}(X||Y) = H_\alpha^R(X+Y) - \frac{H_\alpha^R(X) + H_\alpha^R(Y)}{2} \quad (4.62)$$

definiert und im Folgenden als *modifizierte Jensen-Rényi-Divergenz* bezeichnet.

Für den Fall von  $n$  Verteilungen erhält man dann

$$D_\alpha^{modJR}(X_1||X_2||\dots||X_n) = H_\alpha^R\left(\sum_{i=1}^n X_i\right) - \frac{1}{n} \sum_{i=1}^n H_\alpha^R(X_i). \quad (4.63)$$

Wählt man  $\alpha = 1$ , so wird die modifizierte Jensen-Rényi-Divergenz zur modifizierten Jensen-Shannon-Divergenz (Gleichung (4.49)).

- **Tsallis Entropie.** Die vorige Überlegung lässt sich auch auf die Tsallis Entropie aus der Gleichung (2.48) anwenden. Aus der Gleichung (2.48) folgt

$$H_q^T(X) = \frac{1}{q-1} \left(1 - \|f_X^q(x)\|_q^q\right) \quad (4.64)$$

$$= \frac{1}{1-q} \left(\|f_X^q(x)\|_q^q - 1\right). \quad (4.65)$$

Betrachtet man weiter den Zusammenhang in der Gleichung (4.53) so gilt auch für die Tsallis Entropie

$$H_q^T(X + Y) \propto \|f_X(x) * f_Y(x)\|_q. \quad (4.66)$$

Erweitert man nun die (4.55) und (4.56) auf Tsallis Entropie, so erhält man

$$\frac{1}{1-q} \left( \|f * g\|_q^q - 1 \right) \geq \frac{1}{1-q} \left( \|f_X^q(x)\|_q^q - 1 \right) \Rightarrow \quad (4.67)$$

$$H_q^T(X + Y) \geq H_q^T(X) \quad (4.68)$$

mit der selben Argumentation wie für (4.58). Verfolgt man die Idee weiter, so kann man analog zur modifizierten Jensen-Rényi-Divergenz die modifizierte Jensen-Tsallis-Divergenz für zwei

$$D_q^{modJT}(X||Y) = H_q^T(X + Y) - \frac{H_q^T(X) + H_q^T(Y)}{2} \quad (4.69)$$

bzw. für  $n$  Zufallsvariablen

$$D_q^{modJT}(X_1||X_2||\dots||X_n) = H_q^T \left( \sum_{i=1}^n X_i \right) - \frac{1}{n} \sum_{i=1}^n H_q^T(X_i) \quad (4.70)$$

definieren. Wählt man  $\alpha = 1$ , so wird die modifizierte Jensen-Tsallis-Divergenz zur modifizierten Jensen-Shannon-Divergenz (Gleichung (4.49)).

### 4.3.3. Algorithmen zu Permutationskorrekturen

Ausgehend von den oben definierten Ähnlichkeitsfunktionen kann dann der **Algorithmus 1** zu Permutationskorrekturen benutzt werden. Dieser berechnet die globale Abhängigkeit  $D \left[ f \left( \mathbf{P}(\Omega_k) \mid \hat{\mathbf{Y}}(\Omega_k, \tau) \right), f \left( \left| \hat{\mathbf{Y}}(\Omega_{k+1}, \tau) \right| \right) \right]$  (Gleichung (4.34)) zwischen zwei benachbarten Frequenzbändern.

Allerdings, wie in [SMM03] gezeigt wurde, kann die Verwendung von nur einem Frequenzband als Referenz zu Fehlkorrekturen führen. Um das Problem zu umgehen wurde in [MIZ01] ein Verfahren vorgestellt, das eine Mittelung über mehrere bereits korrigierte Frequenzbänder als Referenz benutzt. Somit wird die Distanz aus der

**Algorithmus 1:** Permutationskorrektur basierend auf der Berechnung der Distanz zwischen den Amplituden der benachbarten Frequenzbänder unter der Annahmen der Weibull-Verteilung.

---

```

begin
1  Initialisierung. Beginnend mit der Frequenz  $k = N_{FFT}/2 + 1$  berechne die
   Parameter der Weibull-Verteilung von  $|\hat{\mathbf{Y}}(\Omega_k, \tau)|$  [Coh65].
2   $k := k - 1$ 
3  while  $k > 0$  do
4      Berechne die Parameter der Weibull-Verteilung von  $|\hat{\mathbf{Y}}(\Omega_k, \tau)|$ .
5      Berechne die Distanz  $D \left[ f \left( |\hat{\mathbf{Y}}^P(\Omega_k, \tau)| \right), f \left( |\hat{\mathbf{Y}}(\Omega_{k+1}, \tau)| \right) \right]$  wie in der
   Gleichung (4.34) für alle möglichen Permutationen von  $|\hat{\mathbf{Y}}(\Omega_k, \tau)|$ .
6      Wähle die Permutation  $\mathbf{P}_+(\Omega_k)$  mit dem optimalen Wert von  $D$ .
7       $k := k - 1$ 
   end
end

```

---

Gleichung (4.34) neu definiert

$$D \left[ f \left( |\hat{\mathbf{Y}}^P(\Omega_k, \tau)| \right), f \left( \frac{1}{L} \sum_{l=k+1}^{k+1+L} |\hat{\mathbf{Y}}(\Omega_l, \tau)| \right) \right] = \sum_{n=1}^N d \left[ f \left( |\hat{\mathbf{Y}}_n^P(\Omega_k, \tau)| \right), f \left( \frac{1}{L} \sum_{l=k+1}^{k+1+L} |\hat{\mathbf{Y}}_n(\Omega_l, \tau)| \right) \right] \quad (4.71)$$

mit  $L$  als der Anzahl der bereits korrigierten Frequenzbänder für die Mittelung. Somit wird der **Algorithmus 1** zu **Algorithmus 2**.

Das selbe Schema kann bei der Verwendung der Logarithmen der Amplituden der getrennten Signale  $\log |\hat{\mathbf{Y}}(\Omega_k, \tau)|$  statt der Amplituden  $|\hat{\mathbf{Y}}(\Omega_k, \tau)|$  benutzt werden. Wie oben schon erwähnt, folgen die Logarithmen der Amplituden der generalisierten Gaußverteilung (GGD). Somit wird der **Algorithmus 2** zu **Algorithmus 3**.

## 4.4. Permutationskorrekturen basierend auf den Kernfunktionen

Die im oberen Abschnitt beschriebenen Verfahren basieren auf der Schätzung von Parametern der angenommenen Verteilungsdichte. Wird diese Annahme falsch getroffen, so kann unter Umständen Fehler in den Berechnungen verursachen. Dies kann

---

**Algorithmus 2:** Permutationskorrektur basierend auf der Berechnung der Abhängigkeit zwischen den Amplituden des aktuellen Frequenzbandes und einer Mittelung von den bereits korrigierten Bändern unter der Annahmen der Weibull-Verteilung.

---

```

begin
1  Initialisierung. Beginnend mit der Frequenz  $k = N_{FFT}/2 + 1$  berechne die
   Parameter der Weibull-Verteilung von  $|\hat{\mathbf{Y}}(\Omega_k, \tau)|$  [Coh65].
2   $k := k - 1$ 
3  while  $k > 0$  do
4  |   Berechne die Parameter der Weibull-Verteilung von  $|\hat{\mathbf{Y}}(\Omega_k, \tau)|$  und von
   |   dem Mittelwert von  $L$  bereits korrigierten Frequenzbände
   |    $\frac{1}{\hat{L}-(k+1)} \sum_{l=k+1}^{\hat{L}} |\hat{\mathbf{Y}}(\Omega_l, \tau)|$ , mit  $\hat{L} = \min(k + 1 + L, N_{FFT}/2 + 1)$ .
5  |   Berechne  $D \left[ f \left( |\hat{\mathbf{Y}}^P(\Omega_k, \tau)| \right), f \left( \frac{1}{\hat{L}} \sum_{l=k+1}^{\hat{L}} |\hat{\mathbf{Y}}(\Omega_l, \tau)| \right) \right]$  für alle möglichen
   |   Permutationen von  $|\hat{\mathbf{Y}}(\Omega_k, \tau)|$ , wie in der Gleichung (4.71) definiert wurde.
6  |   Wähle die Permutation  $\mathbf{P}_+(\Omega_k)$  mit dem optimalen Wert von  $D$ .
7  |    $k := k - 1$ 
   end
end

```

---

z.B. passieren, wenn man statt Sprachsignale (wie dies in der vorliegenden Arbeit angenommen wird) andere akustische Signale trennt. Eine alternative Vorgehensweise wäre es, die Merkmale der getrennten Signale zu untersuchen und auf diese dann gewöhnliche Abstandsmaße anzuwenden. Diese Möglichkeit ist durch die Anwendung des Kernel-Tricks (siehe Abschnitt 2.3) gegeben. Wendet man also auf jedes Paar der getrennten Signale  $|\hat{\mathbf{Y}}(\Omega_k, \tau)|$  und  $\frac{1}{\hat{L}-(k+1)} \sum_{l=k+1}^{\hat{L}} |\hat{\mathbf{Y}}(\Omega_l, \tau)|$ <sup>8</sup> eine Kernfunktion an, so ist die resultierende Größe als ein Skalarprodukt

$$\begin{aligned}
 \kappa \left( \left| \hat{\mathbf{Y}}(\Omega_k, \tau) \right|, \frac{1}{\hat{L} - (k + 1)} \sum_{l=k+1}^{\hat{L}} \left| \hat{\mathbf{Y}}(\Omega_l, \tau) \right| \right) \\
 = \left\langle \Phi \left( \hat{\mathbf{Y}}(\Omega, \tau) \right), \Phi \left( \frac{1}{\hat{L} - (k + 1)} \sum_{l=k+1}^{\hat{L}} \left| \hat{\mathbf{Y}}(\Omega_l, \tau) \right| \right) \right\rangle \quad (4.72)
 \end{aligned}$$

im Merkmalsraum zu interpretieren (siehe Gleichung (2.92)).  $\kappa(\cdot, \cdot)$  bezeichnet dabei eine Kernfunktion, an deren Stelle die Funktionen aus der Tabelle 2.1 eingesetzt

---

<sup>8</sup>Wie in der Gleichung (4.71) bezeichnet  $|\hat{\mathbf{Y}}(\Omega_k, \tau)|$  das aktuelle Frequenzband und  $\frac{1}{\hat{L}-(k+1)} \sum_{l=k+1}^{\hat{L}} |\hat{\mathbf{Y}}(\Omega_l, \tau)|$  den Mittelwert von  $\hat{L}$  bereits korrigierten Frequenzbände, mit  $\hat{L} = \min(k + 1 + L, N_{FFT}/2 + 1)$ .

**Algorithmus 3:** Permutationskorrektur basierend auf der Berechnung der Abhängigkeit zwischen den Logarithmen der Amplituden des aktuellen Frequenzbandes und einer Mittelung der Logarithmen der Amplituden von den bereits korrigierten Bänder unter der Annahmen der Generalisierten Gaussverteilung.

---

**begin**

```

1  Initialisierung. Beginnend mit der Frequenz  $k = N_{FFT}/2 + 1$  berechne die
   Parameter der generalisierten Gaußverteilung von  $\log \left| \hat{\mathbf{Y}}(\Omega_k, \tau) \right|$  (Gleichungen
   (2.20) - (2.24)).
2   $k := k - 1$ 
3  while  $k > 0$  do
4  |   Berechne die Parameter der generalisierten Gaußverteilung von
   |    $\log \left| \hat{\mathbf{Y}}(\Omega_k, \tau) \right|$  und von dem Mittelwert von  $L$  bereits korrigierten
   |   Frequenzbänder  $\log \left( \frac{1}{\hat{L} - (k+1)} \sum_{l=k+1}^{\hat{L}} \left| \hat{\mathbf{Y}}(\Omega_l, \tau) \right| \right)$ , mit
   |    $\hat{L} = \min(k + 1 + L, N_{FFT}/2 + 1)$ .
5  |   Berechne  $D \left[ f \left( \log \left| \hat{\mathbf{Y}}^P(\Omega_k, \tau) \right| \right), f \left( \log \left( \frac{1}{\hat{L}} \sum_{l=k+1}^{\hat{L}} \left| \hat{\mathbf{Y}}(\Omega_l, \tau) \right| \right) \right) \right]$  für alle
   |   möglichen Permutationen von  $\left| \hat{\mathbf{Y}}(\Omega_k, \tau) \right|$ , wie in der Gleichung (4.71)
   |   definiert wurde.
6  |   Wähle die Permutation  $\mathbf{P}_+(\Omega_k)$  mit dem optimalen Wert von  $D$ .
7  |    $k := k - 1$ 
   end
end

```

---

werden können. Ausgehend davon wird an dieser Stelle vorgeschlagen die im Abschnitt 2.3.2 beschriebene Korrentropie (Gleichung (2.99)) als Abstandsmaß für die Permutationskorrekturen zu nehmen [HKO12]. Somit lässt sich eine Kostenfunktion

$$\hat{\mathbf{P}}(\Omega) = \arg \max_{\mathbf{P}(\Omega)} V \left( \left| \hat{\mathbf{Y}}^P(\Omega_k, \tau) \right|, \frac{1}{\hat{L} - (k + 1)} \sum_{l=k+1}^{\hat{L}} \left| \hat{\mathbf{Y}}(\Omega_l, \tau) \right| \right) \quad (4.73)$$

bzw. für die Logarithmen der Amplituden

$$\hat{\mathbf{P}}(\Omega) = \arg \max_{\mathbf{P}(\Omega)} V \left( \log \left| \hat{\mathbf{Y}}^P(\Omega, \tau) \right|, \log \left( \frac{1}{\hat{L} - (k + 1)} \sum_{l=k+1}^{\hat{L}} \left| \hat{\mathbf{Y}}(\Omega_l, \tau) \right| \right) \right) \quad (4.74)$$

aufstellen. Die Korrentropien berechnen sich als

$$\begin{aligned}
 V \left( \left| \hat{\mathbf{Y}}^P(\Omega_k, \tau) \right|, \frac{1}{\hat{L} - (k+1)} \sum_{l=k+1}^{\hat{L}} \left| \hat{\mathbf{Y}}(\Omega_l, \tau) \right| \right) \\
 = E_\tau \left[ \kappa \left( \left| \hat{\mathbf{Y}}^P(\Omega_k, \tau) \right|, \frac{1}{\hat{L} - (k+1)} \sum_{l=k+1}^{\hat{L}} \left| \hat{\mathbf{Y}}(\Omega_l, \tau) \right| \right) \right] \quad (4.75)
 \end{aligned}$$

mit

$$\begin{aligned}
 E_\tau \left[ \kappa \left( \left| \hat{\mathbf{Y}}^P(\Omega_k, \tau) \right|, \frac{1}{\hat{L} - (k+1)} \sum_{l=k+1}^{\hat{L}} \left| \hat{\mathbf{Y}}(\Omega_l, \tau) \right| \right) \right] \\
 \approx \frac{1}{T} \sum_{\tau=1}^T \kappa \left( \left| \hat{\mathbf{Y}}^P(\Omega_k, \tau) \right|, \frac{1}{\hat{L} - (k+1)} \sum_{l=k+1}^{\hat{L}} \left| \hat{\mathbf{Y}}(\Omega_l, \tau) \right| \right). \quad (4.76)
 \end{aligned}$$

$T$  ist dabei die Anzahl der Zeitpunkte, über die gemittelt wird. Damit lässt sich ein weiterer Algorithmus zu Permutationskorrekturen definieren. Dieser behält die vereinfachte Struktur der Algorithmen 2 und 3, da die Parameterschätzung der entsprechenden Verteilungsdichten bei der Berechnung der Korrentropien nach Gleichung (4.76) entfällt.

---

**Algorithmus 4:** Permutationskorrektur basierend auf der Maximierung der Korrentropie zwischen den Amplituden bzw. den Logarithmen der Amplituden des aktuellen Frequenzbandes und einer Mittelung der Amplituden bzw. der Logarithmen der Amplituden von den bereits korrigierten Bändern.

---

```

begin
1    $k = N_{FFT}/2$ 
2   while  $k > 0$  do
3     Berechne  $V \left( \left| \hat{\mathbf{Y}}^P(\Omega_k, \tau) \right|, \frac{1}{\hat{L} - (k+1)} \sum_{l=k+1}^{\hat{L}} \left| \hat{\mathbf{Y}}(\Omega_l, \tau) \right| \right)$  bzw.
        $V \left( \log \left| \hat{\mathbf{Y}}^P(\Omega_k, \tau) \right|, \log \left( \frac{1}{\hat{L} - (k+1)} \sum_{l=k+1}^{\hat{L}} \left| \hat{\mathbf{Y}}(\Omega_l, \tau) \right| \right) \right)$  mit
        $\hat{L} = \min(k + 1 + L, N_{FFT}/2 + 1)$ , für alle möglichen Permutationen von
        $\left| \hat{\mathbf{Y}}(\Omega_k, \tau) \right|$ , mit einer Kernfunktion  $\kappa(\cdot, \cdot)$  aus der Tabelle 2.1.
4     Wähle die Permutation  $\mathbf{P}_+(\Omega_k)$  mit dem maximalen Wert der
       Korrentropie.
5      $k := k - 1$ 
   end
end

```

---



# 5. Verfahren zur Zeit-Frequenz-Maskierung und Spektralen Subtraktion

Im vorigen Kapitel wurde die Vorgehensweise bei der Trennung der konvolutiven Mischungen vorgestellt. Diese beruht auf der Anwendung der instantanen ICA- Algorithmen im Frequenzbereich, was einer linearen Filterung im Zeitbereich entspricht. Jedoch sind die Ergebnisse der Trennung dabei nicht immer befriedigend. Meist sind die Ausgangssignale durch bleibende Übersprecher, Rauschsignale und andere Artefakte gestört, so dass eine Verbesserung der Ergebnisse der Trennung erwünscht ist. Diese ist durch die Verfahren der Zeit-Frequenz-Maskierung gegeben. Die Zeit-Frequenz-Maskierung entspricht dabei einer nichtlinearen zeitvarianten Nachfilterung des Ausgangssignals.

In diesem Kapitel werden die bereits existierenden Verfahren der Zeit-Frequenz-Maskierung diskutiert sowie neue Methoden vorgestellt. Diese basieren zum Teil auf den Methoden der einkanaligen Störgeräuschunterdrückung.

## 5.1. Überblick über die Verfahren der Zeit-Frequenz-Maskierung

Die Idee der Signalverbesserung durch die Maskierung der Zeit-Frequenz-Punkte, an denen ein Störsignal vermutet wird, ist nicht neu und ist auf die ersten Entwicklungen der Verfahren einkanaliger Geräuschreduktion zurückzuführen, solche wie z.B. Wiener-Filter [LO79, Loi07]. Zeit-Frequenz-Maskierungsverfahren, die unter dem Aspekt der Quellentrennung entstanden sind, kann man grob in zwei Bereiche unterteilen: Methoden der *Computational Auditory Scene Analysis* (CASA) und statistischen Methoden, die unter anderem auf der ICA bzw. Beamforming basieren. Bei den CASA-Methoden geht es darum, die akustischen Signale vom

Gesichtspunkt der menschlichen Wahrnehmungs- und Verarbeitungsprinzipien zu betrachten und zu zerlegen [Lyo83, Bre90]. Im Gegensatz dazu gehen die statistischen Verfahren von rein mathematischen Betrachtungen aus und nutzen dabei die Eigenschaften der Disjunktivität der Quellsignale in der Zeit<sup>1</sup>. Weiterhin lassen sich die Zeit-Frequenz-Maskierungsverfahren in einkanalige oder monaurale (hierzu gehören fast alle CASA-Verfahren) und mehrkanalige Methoden unterteilen. Unter den monauralen Methoden kann man wiederum zwei Gruppen hervorheben, nämlich die merkmals- und die modellbasierten Verfahren [WB06]. Die ersten Methoden extrahieren die Maskenfunktion aus der Analyse und Gruppierung bestimmter Merkmale, wie Grundfrequenz und Harmonischen, Offset/Onset, Amplituden- und Frequenzmodulation, zeitliche Kontinuität, Periodizität und zeitlichen Korrelationen [BC94, HW04]. Die modellbasierten Systeme gehen von den trainierten Sprach- und Rauschmodellen aus [Ell06, RR07]. Die Modelle können dabei in Form von Hidden-Markov-Modellen (HMM) [Row00], Gaussian mixture models (GMMs) [RR07] etc. vorliegen. Das Hauptproblem bei den monauralen Maskierungsmethoden ist ein hoher Rechenaufwand, der durch die komplizierten Operationen der Merkmalsextraktion, Gruppierung, Segmentierung oder Modellbildung und Training verursacht wird. Dies macht sie eher vom theoretischen Gesichtspunkt aus interessant und in der Praxis momentan nur bedingt einsetzbar.

Wenn die monauralen Methoden eher als CASA-Methoden einzustufen sind, fallen die mehrkanaligen Verfahren sowohl unter CASA- als auch unter die statistischen Kategorien. Unter den mehrkanaligen Verfahren nehmen die zweikanaligen bzw. die binauralen Methoden eine Sonderstellung ein. Hier wird meistens eine besondere Platzierung und Ausrichtung der Mikrofone vorausgesetzt, die vom potentiellen Einsatz dieser Methoden auf dem Gebiet der Hörhilfen motiviert sind. Unter solchen Verfahren sind besonders die von [RW04, BKP<sup>+</sup>08] hervorzuheben. Dabei ist anzumerken, dass die binauralen Methoden durch getroffene Annahmen und Einschränkungen meistens eine auf das Problem zugeschnittene Lösung liefern und nur schwer bzw. gar nicht erweiterbar sind.

Zur Verdeutlichung ist die oben gegebene Übersicht über die Verfahren der Zeit-Frequenz-Maskierung in Abbildung 5.1 zusammengefasst.

Im Fokus der vorliegenden Arbeit liegen die statistischen Verfahren der Zeit-Frequenz-Maskierung und insbesondere diejenigen auf Basis von ICA. Diese arbeiten auf den Ausgangssignalen von blinder Quellentrennung. Eine Ausnahme stellt

---

<sup>1</sup>Dies ist in der englischsprachigen Literatur als *Sparsity* bekannt.

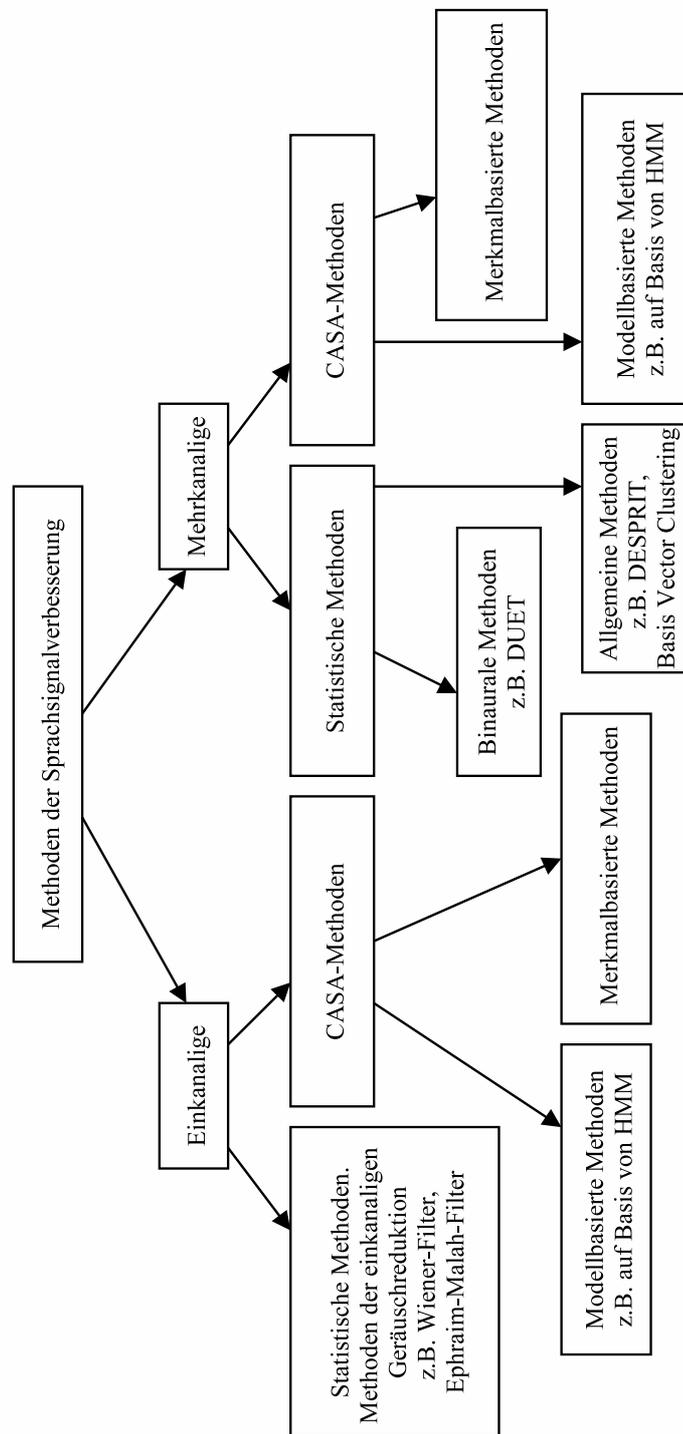


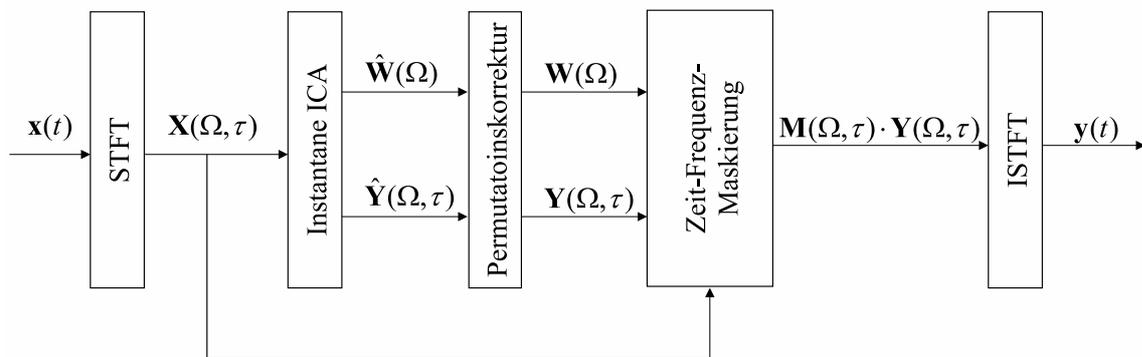
Abbildung 5.1.: Übersicht über die Verfahren der Zeit-Frequenz-Maskierung.

dabei das DUET-Verfahren [YR04] dar, das prinzipiell auf den gemischten Signalen arbeitet, das aber auch für die Benutzung in Verbindung mit ICA erweitert werden kann. Das Ziel der Maskierungsalgorithmen ist es, eine Maske  $\mathbf{M}$  mit  $0 \leq M_i(\Omega, \tau) \leq 1, 1 \leq i \leq N$  zu finden, die bei der Multiplikation mit den durch ICA geschätzten Quellsignalen

$$\tilde{\mathbf{Y}}_i(\Omega, \tau) = M_i(\Omega, \tau)\mathbf{Y}_i(\Omega, \tau) \quad (5.1)$$

die bleibenden Interferenzen minimiert. Der Datenfluss bei der Quellentrennung mit Zeit-Frequenz-Maskierung ist in der Abbildung 5.2 dargestellt.

In den weiteren Abschnitten werden die Maskierungsverfahren genauer vorgestellt, die für die vorliegende Arbeit von Interesse sind.



**Abbildung 5.2.:** Datenfluss bei der frequenzbasierten Entmischung konvolutiver Mischungen mit Zeit-Frequenz-Maskierung.

### 5.1.1. DUET-Algorithmus

Das DUET-Verfahren<sup>2</sup> von Yilmaz und Ricard [YR04] geht von der Annahme aus, dass in einem halltoten Raum pro Zeitfenster  $\tau$  höchstens ein Quellsignal pro Frequenz aktiv ist. Das heißt für zwei unabhängige Signale  $S_i(\Omega, \tau)$  und  $S_j(\Omega, \tau)$

$$S_i(\Omega, \tau)S_j(\Omega, \tau) = 0, \quad \forall(\Omega, \tau), \quad \forall i \neq j \quad (5.2)$$

mit  $1 \leq i \leq N$ ,  $1 \leq j \leq N$  und  $N$  als Anzahl der Quellsignale. Diese Annahme wird auch als Annahme der *W-disjunkten Orthogonalität* (engl. *W-disjoint orthogonality* bzw. *W-DO*<sup>3</sup>) bezeichnet und erlaubt im Idealfall die Extraktion von  $N$  mit  $N \geq 2$  Sprecher aus nur zwei Mikrophonaufnahmen. Die Zeit-Frequenz-Maskierungsfunktion für das  $i$ -te Quellsignal  $S_i(\Omega, \tau)$  lässt sich dann definieren als

$$M_i(\Omega, \tau) = \begin{cases} 1, & S_i(\Omega, \tau) \neq 0 \\ 0, & \text{sonst} \end{cases}. \quad (5.3)$$

Die Bestimmung der Maskenfunktion erfolgt ausgehend von einem normierten Mischungsmodell

$$\begin{bmatrix} X_1(\Omega, \tau) \\ X_2(\Omega, \tau) \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ a_1 e^{-j\Omega\delta_1} & \dots & a_N e^{-j\Omega\delta_N} \end{bmatrix} \begin{bmatrix} S_1(\Omega, \tau) \\ \vdots \\ S_N(\Omega, \tau) \end{bmatrix}, \quad (5.4)$$

wobei  $S_i(\Omega, \tau)$  ein Quellsignal,  $X_1(\Omega, \tau)$  und  $X_2(\Omega, \tau)$  die beiden Mischungen,  $a_n$  die Dämpfungs- und  $\delta_n$  Verzögerungsparameter der Mischung bezeichnen. Dabei geht man davon aus, dass  $a_n$  und  $\delta_n$  frequenzinvariant bleiben [YR04]. Für jeden Punkt  $(\Omega, \tau)$  werden die approximierten Dämpfungs- und Verzögerungsparameter  $\tilde{a}(\Omega, \tau)$  und  $\tilde{\delta}(\Omega, \tau)$

$$\tilde{a}(\Omega, \tau) = |R_{21}(\Omega, \tau)| \quad (5.5)$$

$$\tilde{\delta}(\Omega, \tau) = \frac{1}{\Omega} \angle R_{21}(\Omega, \tau) \quad (5.6)$$

<sup>2</sup>DUET steht für *degenerate unmixing estimation technique*.

<sup>3</sup>W steht in diesem Fall für Fenster (engl. *window*) und deutet auf die Orthogonalität der Signale pro Zeitfenster.

berechnet, mit

$$R_{21}(\Omega, \tau) = \frac{X_2(\Omega, \tau)}{X_1(\Omega, \tau)}. \quad (5.7)$$

Ausgehend vom normierten Mischungsmodell 5.4 und der W-DO Annahme lässt sich leicht schließen, dass für die Zeit-Frequenzpunkte  $(\Omega^i, \tau^i)$ , wenn nur das Signal  $S_i(\Omega, \tau)$  aktiv ist, gilt:

$$\tilde{a}(\Omega^i, \tau^i) = a_i \quad (5.8)$$

und

$$\tilde{\delta}(\Omega^i, \tau^i) = \delta_i. \quad (5.9)$$

Um nun die Dämpfungs- und Verzögerungsparameter der normierten Mischungsmatrix zu schätzen, wird über alle Paare  $\langle \tilde{a}(\Omega, \tau), \tilde{\delta}(\Omega, \tau) \rangle$  ein zwei dimensionales Histogramm aufgespannt, in dem nach dominanten Peaks gesucht wird, die dann den entsprechenden Quellsignalen zugeordnet werden (Abbildung 5.3). Die Maskierungsfunktion für das Signal  $S_i(\Omega, \tau)$  wird dann berechnet als

$$M_i^{\text{DUET}}(\Omega, \tau) \approx \begin{cases} 1, & (\Omega^i, \tau^i), i \approx \arg \max_i L_i(\Omega, \tau) \\ 0, & \text{sonst} \end{cases} \quad (5.10)$$

mit

$$L_i(\Omega, \tau) = p(X_1(\Omega, \tau), X_2(\Omega, \tau) | a_i, \delta_i) \quad (5.11)$$

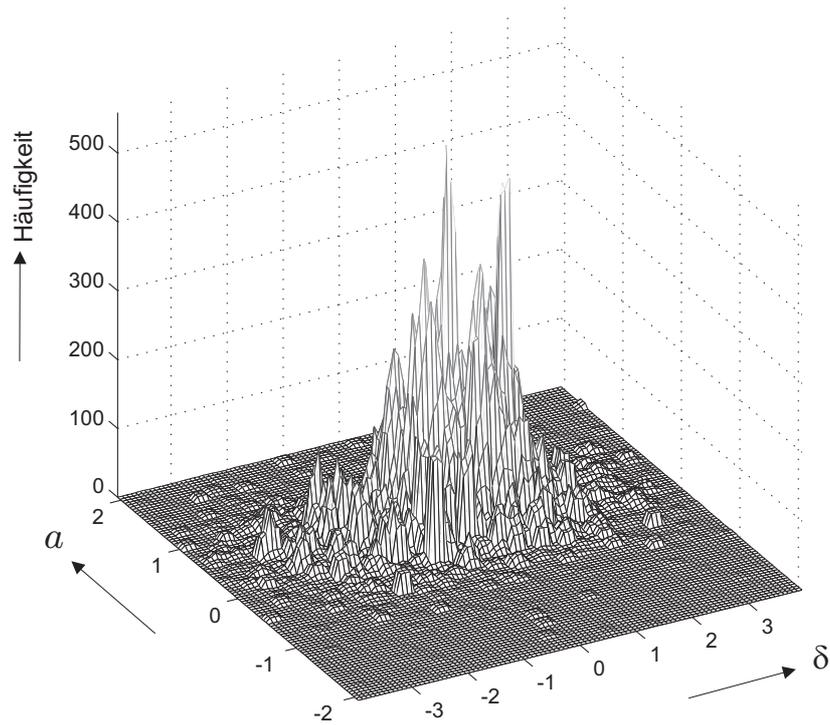
$$= \frac{1}{2\pi\sigma_i^2} \exp \left\{ -\frac{|a_i e^{-j\Omega\delta_i} X_1(\Omega, \tau) - X_2(\Omega, \tau)|^2}{2\sigma_i^2(1 + a_i^2)} \right\} \quad (5.12)$$

als Likelihood-Funktion, wobei  $\sigma_i$  die Varianz der Gaußverteilung darstellt.

### 5.1.2. Amplitudenbasierte Maskierung

Eine weitere Möglichkeit, eine Maskierungsfunktion zu erhalten, ist durch die Schätzung des lokalen Signal-Rausch-Verhältnisses (SNR) aus den getrennten Signalen gegeben [KO09]. Geht man von der Annahme  $Y_i(\Omega, \tau) \approx S_i(\Omega, \tau)$  aus, so ergibt sich die SNR-Schätzung zu

$$\text{SNR}_i = \log_{10} (|Y_i(\Omega, \tau)|^2) - \max_{\forall j \neq i} \log_{10} (|Y_j(\Omega, \tau)|^2), \quad (5.13)$$



**Abbildung 5.3.:** Zweidimensionales Histogramm aufgespannt für alle Paare  $\langle \tilde{a}(\Omega, \tau), \tilde{\delta}(\Omega, \tau) \rangle$ . Das Histogramm wurde für die Aufnahme aus einem realen nicht reverberationsfreien Raum (Reverberationszeit  $T_R = 130$  ms) mit zwei Mikrofonen und vier Sprechern berechnet.

mit  $1 \leq i \leq N$ ,  $1 \leq i \leq N$  und  $N$  als Anzahl der Quellsignale. Die Maske wird dann aus dem Vergleich des lokalen SNRs mit einem Schwellenwert  $T$  gebildet

$$M_i^{\text{Amp}}(\Omega, \tau) = \Psi \left( \text{SNR}_i - \frac{T}{10} \right) \quad (5.14)$$

mit

$$\Psi(x) = \begin{cases} 0, & x \leq 0 \\ 1, & 0 < x. \end{cases} \quad (5.15)$$

### 5.1.3. Phasenbasierte Maskierung

Der Algorithmus von Sawada et al. [SAMM06] beruht, wie der DUET-Algorithmus, auf der W-DO Annahme. Außerdem wird hier die Tatsache ausgenutzt, dass die unabhängigen Quellen im Whitening-Raum orthogonal zueinander sind.

Somit werden bei dieser Methode die gemischten Signale sowie die geschätzten Mischungsmatrizen im Whitening-Raum betrachtet. Der Prozess der Transformation der Signale/Matrizen ist im Abschnitt 3.3 beschrieben und wird durch eine Multiplikation der Signale mit einer Whitening-Matrix

$$\mathbf{U}(\Omega) = \mathbf{\Sigma}^{-\frac{1}{2}} \quad (5.16)$$

vollzogen

$$\mathbf{Z}(\Omega, \tau) = \mathbf{U}(\Omega)\mathbf{X}(\Omega, \tau) \quad (5.17)$$

$$\mathbf{O}(\Omega) = [\mathbf{o}_1(\Omega), \dots, \mathbf{o}_N(\Omega)] = \mathbf{U}(\Omega)\hat{\mathbf{A}}(\Omega) \quad (5.18)$$

mit  $\mathbf{\Sigma} \in \mathbb{C}^{N \times N}$  als Korrelationsmatrix von  $\mathbf{X}(\Omega, \tau)$ , mit  $N$  als Anzahl der Quellsignale(s. Abschnitt 3.3) und  $\mathbf{X}(\Omega, \tau), \mathbf{Z}(\Omega, \tau) \in \mathbb{C}^N$ .

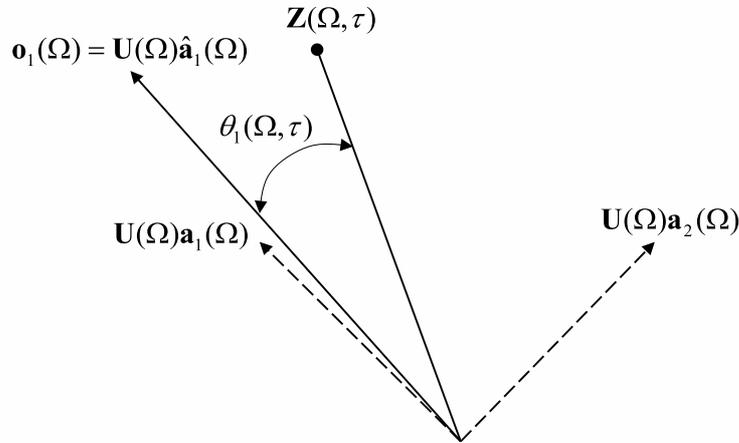
Die Maskenfunktion berechnet sich in Abhängigkeit von den Winkeln zwischen den Vektoren der geschätzten Mischungsmatrix  $\mathbf{O}(\Omega)$  und den dekorellierten Signalen  $\mathbf{Z}(\Omega, \tau)$

$$\theta_i(\Omega, \tau) = \arccos \frac{|\mathbf{o}_i^H(\Omega)\mathbf{Z}(\Omega, \tau)|}{\|\mathbf{o}_i(\Omega)\| \|\mathbf{Z}(\Omega, \tau)\|}. \quad (5.19)$$

Dieser Prozess wird in der Abbildung 5.4 verdeutlicht. Die Maske für das  $i$ -te Quellsignal mit  $1 \leq i \leq N$  und  $N$  als Anzahl der Quellsignale wird dann definiert als

$$M_i^{\text{Phase}}(\theta_i(\Omega, \tau)) = \frac{1}{1 + e^{g(\theta_i(\Omega, \tau) - \theta_T)}}, \quad (5.20)$$

wobei  $\theta_T$  und  $g$  den Schwellwert und die Steilheit der Maske angeben.



**Abbildung 5.4.:** Berechnung der Winkel  $\theta_i(\Omega, \tau)$  im Whitening-Raum für einen Fall von  $N = 2$ . Die Winkel werden zwischen den Vektoren der geschätzten Mischungsmatrix  $\mathbf{O}(\Omega)$  und den dekorelierten Signalen  $\mathbf{Z}(\Omega, \tau)$  berechnet. Die Vektoren  $\mathbf{U}(\Omega)\mathbf{a}_1(\Omega)$  und  $\mathbf{U}(\Omega)\mathbf{a}_2(\Omega)$  verdeutlichen die Lage der tatsächlichen Basisvektoren der Mischungsmatrix.

## 5.2. Interferenzen basierte Maskierung

Im Folgenden wird ein Verfahren zur Zeit-Frequenz-Maskierung vorgestellt, das während der Arbeit an der vorliegenden Dissertation entstand und bei dem eine Maskenfunktion aus den approximierten Maßen für die Dominanz einer extrahierten Quelle bzw. des Störsignals pro Kanal (die so genannten *Interferenzen*) gewonnen wird. Das nachfolgende Verfahren wurde in [HKO07a] publiziert.

### 5.2.1. Extraktion einer Maskenfunktion aus den approximierten Interferenzen

Geht man von der Annahme aus, dass nur eine begrenzte Anzahl der Quellsignale pro Zeit-Frequenz-Schritt dominant ist (die Signale sind *sparse*), und dass die Trennung bis zu einem bestimmten Grad funktioniert hat, so kann man die verbleibenden Interferenzen dadurch detektieren, dass man die Differenz von Nutzsignal und Störsignal zu jedem Zeitpunkt betrachtet. Aus dieser Überlegung heraus kann

man eine Funktion für die Nutzsignalaktivität für die Quelle  $i$

$$U_i(\Omega, \tau, R_\Omega, R_\tau) = \frac{\left\| \Phi(\Omega, \tau, R_\Omega, R_\tau)(Y_i(\Omega, \tau) - \sum_{m \neq i} Y_m(\Omega, \tau)) \right\|}{\left\| \Phi(\Omega, \tau, R_\Omega, R_\tau) \sum_{m \neq i} Y_m(\Omega, \tau) \right\|} \quad (5.21)$$

sowie eine Funktion für die Störsignalinterferenz

$$N_i(\Omega, \tau, R_\Omega, R_\tau) = \frac{\left\| \Phi(\Omega, \tau, R_\Omega, R_\tau)(Y_i(\Omega, \tau) - \sum_{m \neq i} Y_m(\Omega, \tau)) \right\|}{\left\| \Phi(\Omega, \tau, R_\Omega, R_\tau) Y_i(\Omega, \tau) \right\|}. \quad (5.22)$$

definieren. Dabei bezeichnet  $\|\cdot\|$  die Euklidische Norm und

$$\Phi(\Omega, \tau, R_\Omega, R_\tau) = \begin{cases} \mathcal{W}(\Omega - \Omega_0, \tau - \tau_0, R_\Omega, R_\tau), & |\Omega - \Omega_0| \leq R_\Omega, \\ & |\tau - \tau_0| \leq R_\tau \\ 0, & \text{sonst} \end{cases} \quad (5.23)$$

ist eine zweidimensionale Fensterfunktion  $\mathcal{W}(\Omega - \Omega_0, \tau - \tau_0, R_\Omega, R_\tau)$  der Größe  $R_\Omega \times R_\tau$  (z.B. ein zweidimensionales Hanningfenster) und  $1 \leq m \leq N$  und  $N$  als Anzahl der Quellsignale. Die Berechnung von Funktionen 5.21 und 5.22 für einen  $2 \times 2$  Fall ist in der Abbildung 5.5 dargestellt.

Eine Maskierungsfunktion lässt sich nun aus dem Vergleich der Funktionen 5.21 und 5.22 mit einem Schwellenwert definieren. Und zwar ergeben sich mehrere Optionen für eine Maske:

- Die Maskierungsfunktion markiert die Stellen im Signal  $i$ , wo nur das Zielsignal aktiv ist:

$$M_i(\Omega, \tau) = \begin{cases} 1, & U_i(\Omega, \tau, R_\Omega, R_\tau) > \lambda_u \\ 0, & \text{sonst} \end{cases} \quad (5.24)$$

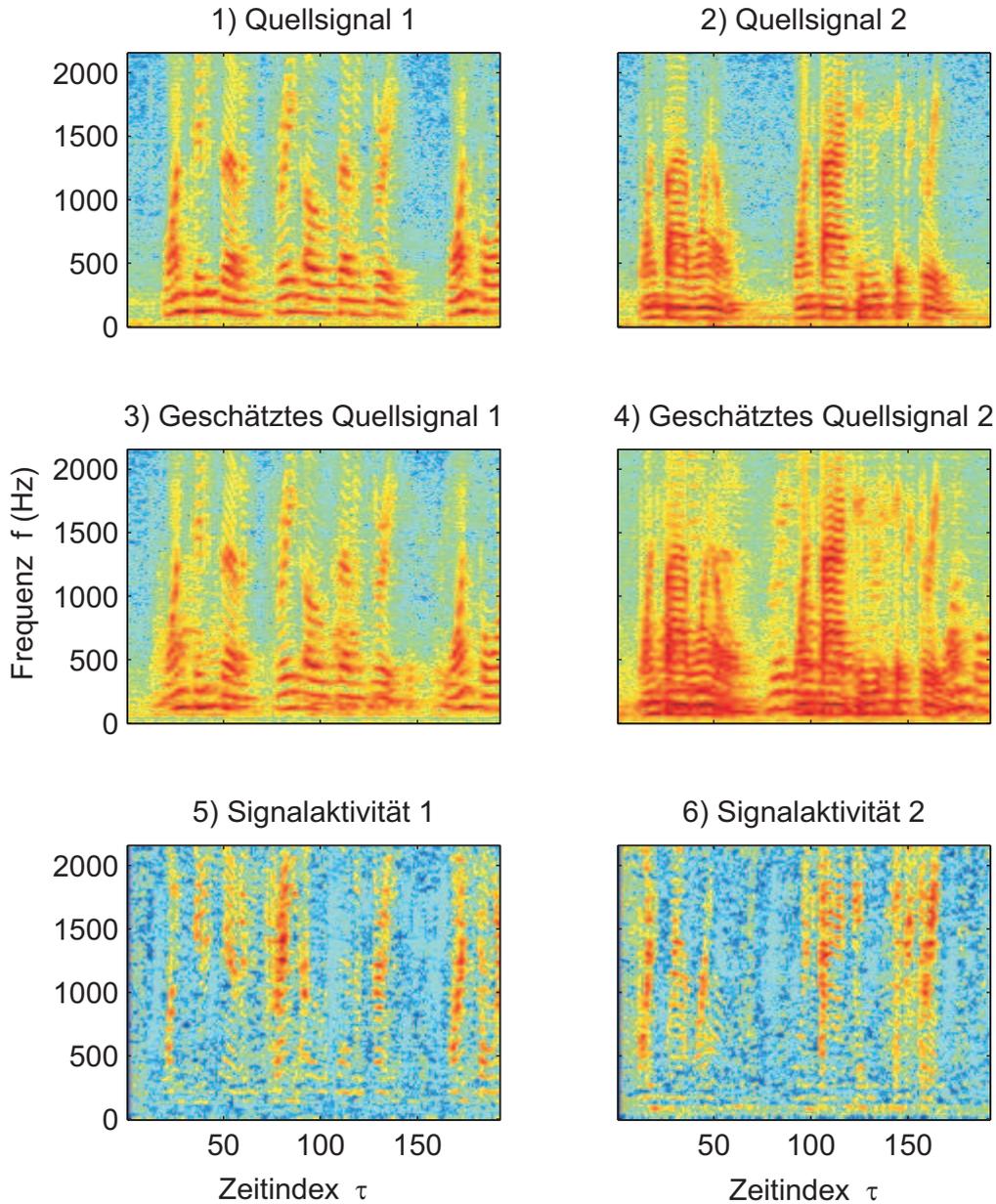
mit  $\lambda_u$  als Schwellenwert.

- Die Maskierungsfunktion setzt die Stellen  $i$  auf 0, wo das Störsignal aktiv ist, oder

$$M_i(\Omega, \tau) = \begin{cases} 0, & N_i(\Omega, \tau, R_\Omega, R_\tau) > \lambda_n \\ 1, & \text{sonst} \end{cases} \quad (5.25)$$

mit  $\lambda_n$  als Schwellenwert für die Störsignalinterferenz.

- Kombination der beiden Methoden:



**Abbildung 5.5.:** Berechnung der Interferenzen für einen  $2 \times 2$  Fall. Die Signalmischung wurde in einem Raum mit einer Reverberationszeit  $T_R = 159$  ms aufgenommen. Die Abbildungen 1) und 2) stellen die unteren 200 Frequenzbänder der Spektrogramme der Quellsignale dar 3) und 4) die entsprechenden ICA-Ausgänge nach der Permutationskorrektur, 5) die Signalaktivität  $U_1(\Omega, \tau, R_\Omega, R_\tau)$  berechnet nach der Gleichung 5.21. 6) die Signalaktivität  $U_2(\Omega, \tau, R_\Omega, R_\tau)$ . Für den  $2 \times 2$  Fall ist  $U_1(\Omega, \tau, R_\Omega, R_\tau) = N_2(\Omega, \tau, R_\Omega, R_\tau)$  und  $U_2(\Omega, \tau, R_\Omega, R_\tau) = N_1(\Omega, \tau, R_\Omega, R_\tau)$  nach der Gleichung 5.22.

$$M_i(\Omega, \tau) = \begin{cases} 1, & N_i(\Omega, \tau, R_\Omega, R_\tau) < \lambda_n \wedge U_i(\Omega, \tau, R_\Omega, R_\tau) > \lambda_u \\ 0, & \text{sonst} \end{cases} \quad (5.26)$$

Bei allen drei Methoden werden binäre Masken berechnet, was unter Umständen zu deutlichen Artefakten im Signal führen kann. Eine Möglichkeit dieses Problemes Herr zu werden, ist durch eine Glättung der Maskierungsfunktion gegeben. Dies wird erreicht, wenn man statt einer binären z.B. die logistische Funktion (wie in Gleichung 5.20) benutzt. In diesem Fall nimmt z.B. die Gleichung 5.26 folgende Form an:

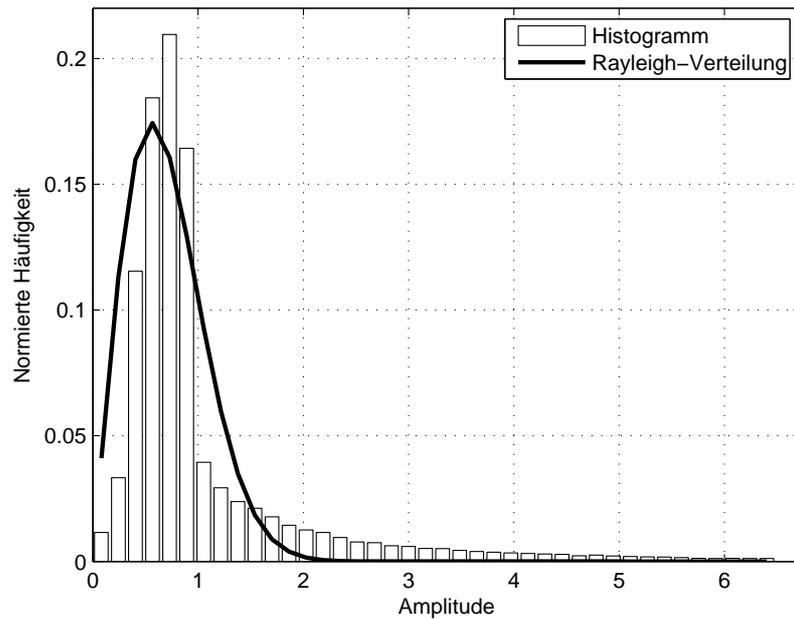
$$M_i^{\text{Int}}(\Omega, \tau) = \frac{1}{1 + \exp(g(U_i(\Omega, \tau) - \lambda_u))} \times \left( 1 - \frac{1}{1 + \exp(g(N_i(\Omega, \tau) - \lambda_n))} \right), \quad (5.27)$$

wobei  $g > 0, g \in \mathbb{R}$  ein Skalierungsfaktor ist. Wie aus den vorigen Gleichungen zu sehen ist, basiert die Berechnung der Maskenfunktion auf den Schwellenwerten für das Nutz- und Störsignal ( $\lambda_u$  und  $\lambda_n$ ). Die Bestimmung der Schwellenwerte ist ein weiteres Problem und wird im folgenden Abschnitt behandelt.

### 5.2.2. Bestimmung der Schwellenwerte

Die Schwellenwerte  $\lambda_u$  und  $\lambda_n$  sind Daten- und Mischungsart abhängig. Somit müssen diese für jede neue Trennung neu berechnet werden. Die Bestimmung der Schwellenwerte erfolgt in Abhängigkeit von den statistischen Eigenschaften der Nutzsignalaktivität und Störsignalinterferenz. Dazu werden erst deren Verteilungsdichten in Betracht gezogen. Da sowohl Nutzsignalaktivität als auch Störsignalinterferenz stets positiv und reellwertig sind (Gl. 5.21-5.22) und aus den Amplituden der getrennten Signale berechnet werden, können sie näherungsweise mit Hilfe der Rayleigh-Verteilung dargestellt werden. Die Abbildung 5.6 zeigt eine Gegenüberstellung von einem Histogramm einer Nutzsignalaktivität und der Rayleigh-Verteilung, die für die vorliegende Nutzsignalaktivität berechnet wurde.

Die kumulative Verteilungsfunktion einer Verteilung ist durch die Gleichung 2.5 gegeben. Für die Rayleigh-Verteilung nimmt die Gleichung folgende



**Abbildung 5.6.:** Gegenüberstellung der wahren Verteilungsdichte einer Nutzsignalaktivität von einem durch ICA getrennten Signal (Gl. 5.21) und einer Rayleigh-Verteilung (schwarze Kurve). Das Sprachsignal wurde in einem Raum mit einer Reverberationszeit  $T_R = 159$  ms aufgenommen. Für den Vergleich wurden 10 Sekunden der Aufnahme verwendet.

Form

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad (5.28)$$

$$= \int_{-\infty}^x \frac{t}{\sigma_X^2} \exp\left\{-\frac{t^2}{2\sigma_X^2}\right\} dt \quad (5.29)$$

$$= 1 - \exp\left\{-\frac{x^2}{2\sigma_X^2}\right\} \quad (5.30)$$

an, wobei  $\sigma_X$  die Standardabweichung von  $X$  ist. Geht man von dieser Formel aus, so stellt sich die Frage, wie man einen Schwellenwert  $\lambda_x$  wählen muss, so dass die Werte der Zufallsvariablen  $X$ , die darunter liegen, mit der Wahrscheinlichkeit  $P(X \leq \lambda_x)$  als **nicht** Zielsignal (bzw. als nicht Störer im Falle der Störsignalinterferenz) zu klassifizieren sind. D.h.

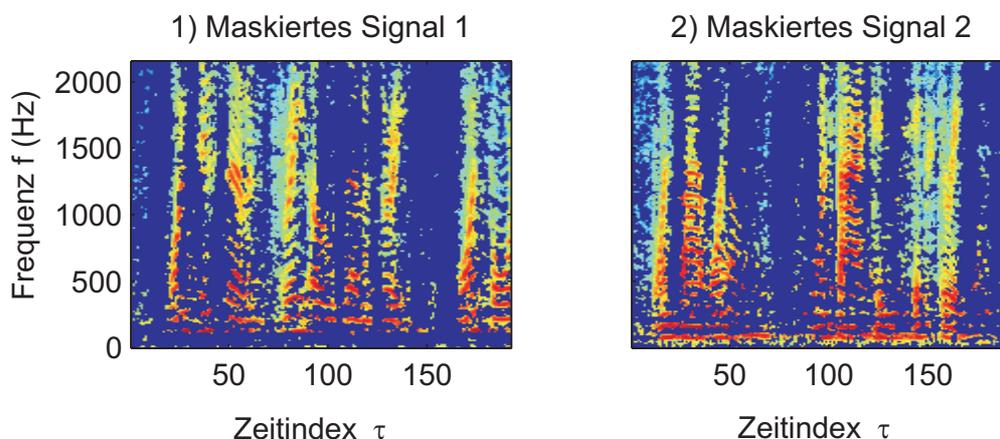
$$P(X \leq \lambda_x) = F_X(\lambda_x) = 1 - \exp\left\{-\frac{\lambda_x^2}{2\sigma_X^2}\right\}. \quad (5.31)$$

Wählt man  $P(X \leq \lambda_x) = P_\lambda$  und stellt die vorige Gleichung nach  $\lambda_x$  um, so erhält

man einen Ausdruck für die Bestimmung der Schwellenwerte:

$$\lambda_x = \sqrt{2\sigma_x^2 \ln \frac{1}{1 - P_\lambda}}. \quad (5.32)$$

Die Abbildung 5.7 zeigt zwei durch die Zeit-Frequenz-Maskierung gewonnene Signale im Frequenzbereich. Hier wurde die Maske nach der Gleichung 5.26 berechnet und auf die Signale aus der Abbildung 5.5 angewandt. Die Berechnung der Schwellenwerte  $\lambda_u$  und  $\lambda_n$  erfolgte nach der Gleichung 5.32 mit  $P_\lambda = 0,5$  für beide Schwellenwerte.



**Abbildung 5.7.:** Maskierte Signale für einen  $2 \times 2$ -Fall. Die Signale wurden auf der gleichen Basis aufgenommen wie in Abbildung 5.5. Die Abbildungen 1) und 2) stellen die unteren 200 Frequenzbänder der Spektrogramme von Signalen dar, die mit Hilfe der Maske aus der Gleichung 5.26 berechnet wurden. Die Berechnung der Schwellenwerte  $\lambda_u$  und  $\lambda_n$  erfolgte nach der Gleichung 5.32 mit  $P_\lambda = 0,5$  für beide Schwellenwerte.

### 5.3. Verfahren der einkanaligen Störgeräuschreduktion

In den vorherigen Abschnitten werden die Methoden der Zeit-Frequenz-Maskierung diskutiert. Handelt es sich dabei um Sprachsignale, so können bei diesen durch die Zeit-Frequenz-Maskierung unerwünschte Artefakte auftreten, die entweder die Sprachverständlichkeit, wenn es sich z.B. um die Spracherkennung handelt, oder den hörbaren Eindruck der Signale deutlich reduzieren oder unter Umständen gar zerstören [HK011]. Es existiert andererseits eine Reihe an Verfahren zur Sprachverbesserung bzw. zur Störgeräuschreduktion in den Sprachsignalen. Diese werden

meistens in einem einkanaligen Szenario eingesetzt und basieren aus diesem Grund auf den A-Priori-Kenntnissen über die statistischen Eigenschaften von Sprache bzw. von Störgeräuschen.

### 5.3.1. Grundlagen der Störgeräuschreduktion

Der Ausgangspunkt für die Verfahren der Störgeräuschreduktion ist das Signalmodell, das eine einkanalige Mikrofonaufnahme in einer nicht rauschfreien Umgebung beschreibt. Dabei geht man von einer additiven Überlagerung von Sprachsignal  $x(t)$  und einer Störung  $d(t)$  aus:

$$y(t) = x(t) + d(t). \quad (5.33)$$

Das Ziel der Störgeräuschreduktion ist nun durch linearphasige zeitvariante Filterung eine Schätzung  $\hat{x}(t)$  des ungestörten Sprachsignals aus der Mikrofonaufnahme  $y(t)$  zu gewinnen, die im Idealfall das gleiche Autoleistungsdichtespektrum wie das Originalsignal aufweist. Die Verfahren für die Lösung dieses Problems kann man grob in drei Klassen unterteilen [Köh05, Loi07]:

- *modellbasierte Rauschunterdrückung* basiert auf dem Vorwissen in Form eines stochastischen Sprachmodells (z.B. Hidden-Markov-Modelle [EMJ89], Kalman-Filter [KGG89], AR-Sprachmodelle [YFD99] etc.) und in manchen Fällen einem Rauschmodell,
- *Subspace-Methoden* zur Rauschunterdrückung zerlegen das rauschbehaftete Signal in die komplementäre Nutzsignal- und Rauschunterräume, meistens auf der Basis von Eigenwertzerlegung der Autokorrelationsmatrix des verrauschten Signals [DBC91, EVT95] und
- Methoden der *Spektralen Subtraktion*.

Für die vorliegende Arbeit sind die Verfahren der Spektralen Subtraktion von Bedeutung. Diese werden im folgenden genauer betrachtet.

Bei der Spektralen Subtraktion handelt es sich um eine Klasse von Verfahren, die die Schätzung des rauschfreien Signals  $\hat{X}(\Omega, \tau)$  aus dem folgenden Zusammenhang

$$\hat{X}(\Omega, \tau) = G(\Omega, \tau)Y(\Omega, \tau) \quad (5.34)$$

gewinnen. Dabei ist  $Y(\Omega, \tau)$  die Kurzzeit-Fourier-Transformierte des Mikrofonsignals und  $G(\Omega, \tau)$  ist die zeit-frequenzabhängige Übertragungsfunktion. Ausschlaggebend für diese Vorgehensweise ist die Äquivalenz von Zeitbereichs- und Fourierbereichsdarstellung von der Gleichung 5.33

$$y(t) = x(t) + d(t) \quad \circ \text{---} \bullet \quad Y(\Omega, \tau) = X(\Omega, \tau) + D(\Omega, \tau). \quad (5.35)$$

Die prinzipielle Arbeitsweise der Spektralen Subtraktion ist in der Abbildung 5.8 veranschaulicht. Wie man aus dem Diagramm leicht erkennen kann bleibt die Phase des zu verarbeitenden Signals dabei unverändert. Die Tatsache, dass die Phase des rauschbefreiten Signals der Phase der Mikrofonaufnahme gleich bleibt, ist dadurch gerechtfertigt, dass das menschliche Gehör wesentlich unempfindlicher auf die Phasenstörungen reagiert, als auf die Störungen in den Amplituden. Dies ist durch mehrere experimentelle Studien belegt worden [WL82].

Ein weiterer wichtiger Punkt bei der Berechnung der Übertragungsfunktion  $G(\Omega, \tau)$  ist die Schätzung der Rauschleistung. Da dies große Auswirkungen auf die Qualität der Spektralen Subtraktion hat, wird dieser Punkt später separat behandelt.

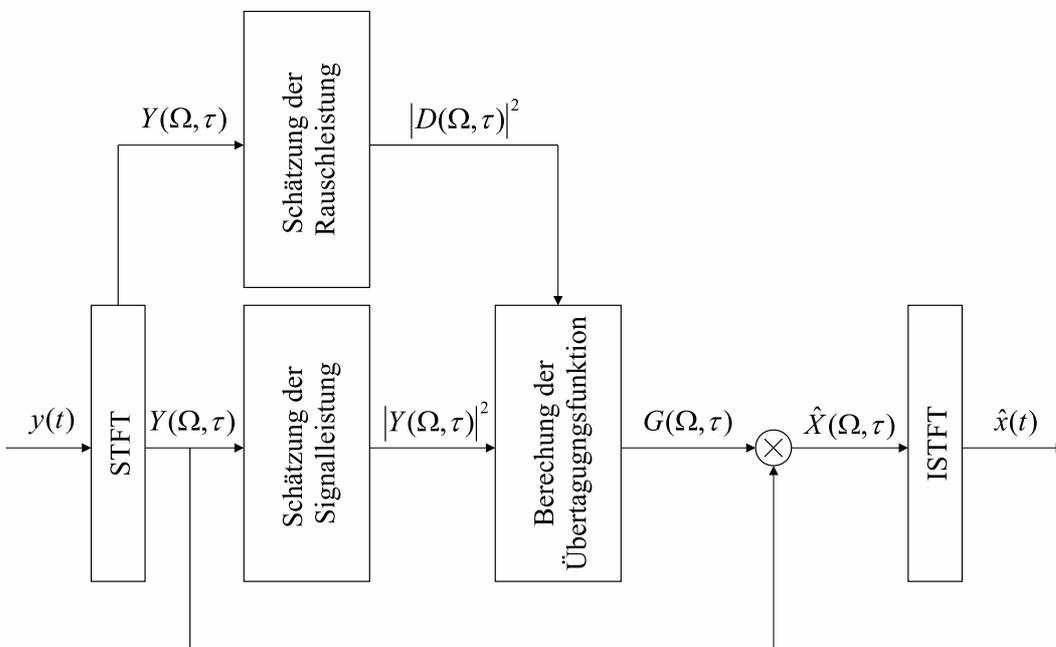


Abbildung 5.8.: Datenfluss bei der Spektralen Subtraktion

### 5.3.2. Verfahren der Spektralen Subtraktion

Im Folgenden werden einige ausgewählte Verfahren der Spektralen Subtraktion im Detail behandelt.

#### 5.3.2.1. Wiener-Filter

Einer der ältesten Ansätze für die Störgeräuschreduktion mit Spektralen Subtraktion ist die Wiener-Filterung [LO79]. Beim Wiener-Filter wird der mittlere quadratische Fehler  $|\mathcal{E}(\Omega, \tau)|^2$

$$E [|\mathcal{E}(\Omega, \tau)|^2] = E \left[ \left| X(\Omega, \tau) - \hat{X}(\Omega, \tau) \right|^2 \right] \stackrel{!}{=} \min \quad (5.36)$$

zwischen dem Originalsignal  $X(\Omega, \tau)$  und seiner Approximation  $\hat{X}(\Omega, \tau)$  minimiert. Dies erfolgt über die Minimierung der Fehlerleistung

$$E [|\mathcal{E}(\Omega, \tau)|^2] = E \left[ \left\{ X(\Omega, \tau) - \hat{X}(\Omega, \tau) \right\}^* \left\{ X(\Omega, \tau) - \hat{X}(\Omega, \tau) \right\} \right] \quad (5.37)$$

$$= E \left[ \left\{ X(\Omega, \tau) - G(\Omega, \tau)Y(\Omega, \tau) \right\}^* \left\{ X(\Omega, \tau) - G(\Omega, \tau)Y(\Omega, \tau) \right\} \right] \stackrel{!}{=} \min, \quad (5.38)$$

die nach der konjugiert Komplexen der Übertragungsfunktion  $G^*(\Omega, \tau)$  abgeleitet wird

$$\frac{\partial E [|\mathcal{E}(\Omega, \tau)|^2]}{\partial G^*(\Omega, \tau)} = G(\Omega, \tau)E [ |Y(\Omega, \tau)|^2 ] - E [ X(\Omega, \tau)Y^*(\Omega, \tau) ] \stackrel{!}{=} 0. \quad (5.39)$$

Stellt man die Gleichung 5.39 nach  $G(\Omega, \tau)$  um, so erhält man die Übertragungsfunktion des entsprechenden Wiener-Filters

$$G(\Omega, \tau) = \frac{E [X(\Omega, \tau)Y^*(\Omega, \tau)]}{E [ |Y(\Omega, \tau)|^2 ]}. \quad (5.40)$$

Geht man von der Annahme aus, dass sowohl das Sprach- als auch das Rauschsignal als unabhängige Gaussvariablen angesehen werden können, so gilt für ihre Varianzen  $\sigma_Y^2(\Omega, \tau)$ ,  $\sigma_X^2(\Omega, \tau)$  und  $\sigma_D^2(\Omega, \tau)$  folgender Zusammenhang

$$E [ |Y(\Omega, \tau)|^2 ] = \sigma_Y^2(\Omega, \tau) = \sigma_X^2(\Omega, \tau) + \sigma_D^2(\Omega, \tau). \quad (5.41)$$

Damit erhält man

$$G(\Omega, \tau) = \frac{E[X(\Omega, \tau)Y^*(\Omega, \tau)]}{\sigma_X^2(\Omega, \tau) + \sigma_D^2(\Omega, \tau)} \quad (5.42)$$

$$= \frac{E[X(\Omega, \tau)(X^*(\Omega, \tau) + D^*(\Omega, \tau))]}{\sigma_X^2(\Omega, \tau) + \sigma_D^2(\Omega, \tau)} \quad (5.43)$$

$$= \frac{E[|X(\Omega, \tau)|^2] + E[X(\Omega, \tau)D^*(\Omega, \tau)]}{\sigma_X^2(\Omega, \tau) + \sigma_D^2(\Omega, \tau)} \quad (5.44)$$

$$= \frac{\sigma_X^2(\Omega, \tau) + E[X(\Omega, \tau)D^*(\Omega, \tau)]}{\sigma_X^2(\Omega, \tau) + \sigma_D^2(\Omega, \tau)}. \quad (5.45)$$

Aufgrund der Annahme der Unabhängigkeit gilt

$$E[X(\Omega, \tau)D^*(\Omega, \tau)] = 0. \quad (5.46)$$

Damit ergibt sich die Übertragungsfunktion zu

$$G^W(\Omega, \tau) = \frac{\sigma_X^2(\Omega, \tau)}{\sigma_X^2(\Omega, \tau) + \sigma_D^2(\Omega, \tau)} \quad (5.47)$$

bzw.

$$G^W(\Omega, \tau) = \frac{\xi(\Omega, \tau)}{\xi(\Omega, \tau) + 1} \quad (5.48)$$

mit

$$\xi(\Omega, \tau) = \frac{\sigma_X^2(\Omega, \tau)}{\sigma_D^2(\Omega, \tau)} \quad (5.49)$$

als *A-Priori*-SNR [Loi07]. Da die A-Priori-Parameter des Sprachsignals meistens nicht zu Verfügung stehen, muss  $\xi(\Omega, \tau)$  geschätzt werden. Auf dieses Problem wird später genauer eingegangen. Alternativ dazu kann die Gleichung 5.40 durch ähnliche Überlegungen wie oben in die Form

$$G(\Omega, \tau) = \frac{E[X(\Omega, \tau)Y^*(\Omega, \tau)]}{\sigma_Y^2(\Omega, \tau)} \quad (5.50)$$

$$= \frac{E[(Y(\Omega, \tau) - D(\Omega, \tau))Y^*(\Omega, \tau)]}{\sigma_Y^2(\Omega, \tau)} \quad (5.51)$$

$$= \frac{E[Y(\Omega, \tau)Y^*(\Omega, \tau)] - E[D(\Omega, \tau)(X^*(\Omega, \tau) + D^*(\Omega, \tau))]}{\sigma_Y^2(\Omega, \tau)} \quad (5.52)$$

$$= \frac{\sigma_Y^2(\Omega, \tau) - \sigma_D^2(\Omega, \tau)}{\sigma_Y^2(\Omega, \tau)} \quad (5.53)$$

gebracht werden. Da sich aber das Sprachsignal nur als kurzzeitig stationäres Signal modellieren lässt, kann die Übertragungsfunktion des Filters nur für einen Signalausschnitt zuverlässig bestimmt werden [MM80]. So wird an Stelle von  $\sigma_Y^2(\Omega, \tau)$  oft  $|Y(\Omega, \tau)|^2$  eingesetzt

$$G(\Omega, \tau) \approx \frac{|Y(\Omega, \tau)|^2 - \sigma_D^2(\Omega, \tau)}{|Y(\Omega, \tau)|^2} \quad (5.54)$$

$$= \frac{\gamma(\Omega, \tau) - 1}{\gamma(\Omega, \tau)} \quad (5.55)$$

mit

$$\gamma(\Omega, \tau) = \frac{|Y(\Omega, \tau)|^2}{\sigma_D^2(\Omega, \tau)} \quad (5.56)$$

als *a-posteriori*-SNR.

Es ist anzumerken, dass die Gleichung 5.48 die gängigere Form des Wiener-Filters darstellt. Diese wurde in [Loi07] auf die parametrische Wiener-Filter Funktion

$$G^W(\Omega, \tau) = \left( \frac{\xi(\Omega, \tau)}{\xi(\Omega, \tau) + a} \right)^\beta \quad (5.57)$$

erweitert. Hierbei sind  $a, \beta > 0$  die Parameter, die die Steilheit des Filters bestimmen.

### 5.3.2.2. Maximum-Likelihood-Subtraktion

Die Berechnung des Wiener-Filters basiert auf der Annahme, dass sowohl das Sprachsignal als auch Rauschen als unabhängige gaußverteilte Zufallsvariablen angesehen werden. In [MM80] wurde ein anderes Modell vorausgesetzt, bei dem das Rauschen als gaußverteilt und die Sprache als deterministische Wellenform von unbekannter Amplitude und Phase angenommen werden. Ausgehend von der Annahme des Gauß-Rauschens wird die Likelihood-Funktion  $f(Y(\Omega, \tau) | X(\Omega, \tau))$  gleich der Verteilungsdichte des Rauschsignals mit dem Mittelwert  $X(\Omega, \tau)$ , was der Gaußverteilung

$$f(Y(\Omega, \tau) | X(\Omega, \tau)) = \frac{1}{\pi\sigma_D^2} \exp\left(-\frac{|Y(\Omega, \tau) - X(\Omega, \tau)|^2}{\sigma_D^2}\right) \quad (5.58)$$

$$= \frac{1}{\pi\sigma_D^2} \exp\left(-\frac{|D(\Omega, \tau)|^2}{\sigma_D^2}\right) \quad (5.59)$$

entspricht. Mit

$$Y(\Omega, \tau) = R(\Omega, \tau) \exp\{j\vartheta(\Omega, \tau)\} \quad (5.60)$$

und

$$X(\Omega, \tau) = A(\Omega, \tau) \exp\{j\varphi(\Omega, \tau)\} \quad (5.61)$$

wird die Gleichung 5.59 zu

$$\begin{aligned} & f(Y(\Omega, \tau) | A(\Omega, \tau), \varphi(\Omega, \tau)) \\ &= \frac{1}{\pi\sigma_D^2} \exp\left(-\frac{R^2(\Omega, \tau) - 2A(\Omega, \tau)\Re\{Y(\Omega, \tau)e^{-j\varphi(\Omega, \tau)}\} + A^2(\Omega, \tau)}{\sigma_D^2}\right), \end{aligned} \quad (5.62)$$

wobei  $\Re\{Y(\Omega, \tau)e^{-j\varphi(\Omega, \tau)}\}$  den Realteil des Ausdruckes  $Y(\Omega, \tau)e^{-j\varphi(\Omega, \tau)}$  bezeichnet. Um nun die Maximum-Likelihood-Schätzung  $\hat{A}(\Omega, \tau)$  vom Amplitudenverlauf des Nutzsignals  $A(\Omega, \tau)$  zu finden, wird die Gleichung 5.62 maximiert

$$\hat{A}(\Omega, \tau) \stackrel{!}{=} \arg \max_{A(\Omega, \tau)} f(Y(\Omega, \tau) | A(\Omega, \tau), \varphi(\Omega, \tau)). \quad (5.63)$$

Da der Phasenverlauf  $\varphi(\Omega, \tau)$  unbekannt und bei der Bestimmung von  $\hat{A}(\Omega, \tau)$  nicht von Bedeutung ist, wird er eliminiert, indem man die Gleichung 5.62 über die Phase mittelt:

$$\overline{f(Y(\Omega, \tau) | A(\Omega, \tau))} = \int_0^{2\pi} f(Y(\Omega, \tau) | A(\Omega, \tau), \varphi(\Omega, \tau)) f(\varphi(\Omega, \tau)) d\varphi(\Omega, \tau). \quad (5.64)$$

Da die wahre Verteilungsdichte von  $\varphi(\Omega, \tau)$  unbekannt ist, wird an dieser Stelle eine Gleichverteilung von  $\varphi(\Omega, \tau)$  auf dem Intervall  $[0, 2\pi]$  angenommen, d.h.  $f(\varphi(\Omega, \tau)) = 1/2\pi, \varphi(\Omega, \tau) \in [0, 2\pi]$ . Dann wird 5.64 zu

$$\begin{aligned} \overline{f(Y(\Omega, \tau) | A(\Omega, \tau))} &= \frac{1}{\pi\sigma_D^2} \exp\left(\frac{R^2(\Omega, \tau) + A^2(\Omega, \tau)}{\sigma_D^2}\right) \\ &\cdot \frac{1}{2\pi} \int_0^{2\pi} \exp\left(-\frac{2A(\Omega, \tau)\Re\{Y(\Omega, \tau)e^{-j\varphi(\Omega, \tau)}\}}{\sigma_D^2}\right) d\varphi(\Omega, \tau). \end{aligned} \quad (5.65)$$

Das Integral in der Gleichung 5.65 ist die modifizierte Besselfunktion erster Ordnung

$$I_0(|x|) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\Re \{e^{-j\varphi} x\}) d\varphi. \quad (5.66)$$

Damit entspricht die Gleichung 5.65 der Rice-Verteilung (Abschnitt 2.1.4.6). Für große Werte von  $x$  ( $x \geq 3$ ) kann die Besselfunktion durch

$$I_0(|x|) \approx \frac{1}{\sqrt{2\pi|x|}} \exp(|x|) \quad (5.67)$$

approximiert werden. Damit wird die Gleichung 5.65 zu

$$\overline{f(Y(\Omega, \tau) | A(\Omega, \tau))} = \frac{1}{\pi\sigma_D^2} \frac{1}{\sqrt{4\pi \frac{A(\Omega, \tau)R(\Omega, \tau)}{\sigma_D^2}}} \exp\left(-\frac{R^2(\Omega, \tau) - 2A(\Omega, \tau)R(\Omega, \tau) + A^2(\Omega, \tau)}{\sigma_D^2}\right). \quad (5.68)$$

Nach dem Differenzieren der log-Likelihood-Funktion  $\log \overline{f(Y(\Omega, \tau) | A(\Omega, \tau))}$  nach  $A(\Omega, \tau)$  und Nullsetzen der Ableitung, bekommt man

$$\hat{A}(\Omega, \tau) = \frac{1}{2} \left[ R(\Omega, \tau) + \sqrt{R^2(\Omega, \tau) - \sigma_D^2(\Omega, \tau)} \right] \quad (5.69)$$

als Schätzwert für  $A(\Omega, \tau)$ . Die Schätzung des unverrauschten Sprachsignals ergibt sich aus der Schätzung für  $A(\Omega, \tau)$  und dem Phasenverlauf des verrauschten Signals  $\varphi(\Omega, \tau)$

$$\hat{X}(\Omega, \tau) = \hat{A}(\Omega, \tau) e^{j\varphi(\Omega, \tau)} \quad (5.70)$$

$$= \hat{A}(\Omega, \tau) \frac{Y(\Omega, \tau)}{R(\Omega, \tau)} \quad (5.71)$$

$$= G^{ML}(\Omega, \tau) Y(\Omega, \tau) \quad (5.72)$$

mit

$$G^{ML}(\Omega, \tau) = \frac{1}{2} + \frac{1}{2} \sqrt{\frac{Y^2(\Omega, \tau) - \sigma_D^2(\Omega, \tau)}{Y^2(\Omega, \tau)}} \quad (5.73)$$

bzw.

$$G^{ML}(\Omega, \tau) = \frac{1}{2} + \frac{1}{2} \sqrt{\frac{\gamma(\Omega, \tau) - 1}{\gamma(\Omega, \tau)}}. \quad (5.74)$$

Mit  $\gamma(\Omega, \tau)$  als a-posteriori-SNR aus der Gleichung 5.56.

In [SF96] wurde die letzte Gleichung auf die Form

$$G^{ML}(\Omega, \tau) = \frac{1}{2} + \frac{1}{2} \sqrt{\frac{\xi(\Omega, \tau)}{\xi(\Omega, \tau) + 1}} \quad (5.75)$$

erweitert. Dabei ist  $\xi(\Omega, \tau)$  die A-Priori-SNR aus der Gleichung 5.49.

### 5.3.2.3. Ephraim-Malah-Filter

Der Ansatz von Ephraim und Malah [EM84], auch bekannt als *Minimum Mean-Square Error Short-Time Spectral Amplitude (MMSE STSA)*, stellt eine Alternative zum Wiener-Filter zur Berechnung der Gewichtskoeffizienten der Spektralen Subtraktion dar. Als einen schönen Nebeneffekt erhält man dabei eine natürlich klingende Reststörung nahezu ohne Verzerrungsartefakte (wie z.B. *Musical Tones*), die bei den „klassischen“ Verfahren durch die unterschiedlichen Varianzen der geschätzten Leistungsdichtespektren von Geräuschen und Sprache entstehen.

Dem Ephraim-Malah-Filter liegt die Annahme zugrunde, dass Real- und Imaginärteile der Fourierkoeffizienten sowohl der Sprache als auch des Rauschens statistisch unabhängig und normalverteilt sind. Dies führt dazu, dass sich die Rayleigh-Verteilung für die Beschreibung der Beträge ergibt [vgl. Abschnitt 4.3.1]. Die Schätzung des Amplitudenverlaufes des reinen Signals kann dann durch die Minimierung des mittleren quadratischen Fehlers der Beträge

$$E \left[ \left( |X(\Omega, \tau)| - |\hat{X}(\Omega, \tau)| \right)^2 \right] = E \left[ \left( A(\Omega, \tau) - \hat{A}(\Omega, \tau) \right)^2 \right] \stackrel{!}{=} \min \quad (5.76)$$

bestimmt werden. Hierbei ist<sup>4</sup>

$$\begin{aligned} E \left[ \left( A(\Omega, \tau) - \hat{A}(\Omega, \tau) \right)^2 \right] \\ = \int_0^\infty \int_0^\infty \left( a(\Omega, \tau) - \hat{A}(\Omega, \tau) \right)^2 f(a(\Omega, \tau), Y(\Omega, \tau)) da(\Omega, \tau) dY(\Omega, \tau). \end{aligned} \quad (5.77)$$

---

<sup>4</sup>Hierbei steht  $a(\Omega, \tau)$  für eine Laufvariable der Zufallsgröße  $A(\Omega, \tau)$ , wie im Abschnitt 2.1.1.

Die Minimierung der Gleichung 5.77 ergibt für den Schätzer der Sprachsignalamplitude den bedingten Erwartungswert ([Tre68], Seite 56, Gleichungen 124-128)

$$\hat{A}(\Omega, \tau) = E[A(\Omega, \tau) | Y(\Omega, \tau)] \quad (5.78)$$

$$= \int_0^{\infty} a(\Omega, \tau) f(a(\Omega, \tau) | Y(\Omega, \tau)) da(\Omega, \tau). \quad (5.79)$$

Für  $f(a(\Omega, \tau) | Y(\Omega, \tau))$  ergibt sich nach dem Bayestheorem

$$f(a(\Omega, \tau) | Y(\Omega, \tau)) = \frac{f(Y(\Omega, \tau) | a(\Omega, \tau)) f(a(\Omega, \tau))}{f(Y(\Omega, \tau))} \quad (5.80)$$

mit

$$\begin{aligned} & f(Y(\Omega, \tau)) \\ &= \int_0^{\infty} \int_0^{2\pi} f(Y(\Omega, \tau) | a(\Omega, \tau), \varphi(\Omega, \tau)) f(a(\Omega, \tau), \varphi(\Omega, \tau)) d\varphi(\Omega, \tau) da(\Omega, \tau) \end{aligned} \quad (5.81)$$

und

$$\begin{aligned} & f(Y(\Omega, \tau) | a(\Omega, \tau)) f(a(\Omega, \tau)) \\ &= \int_0^{2\pi} f(Y(\Omega, \tau) | a(\Omega, \tau), \varphi(\Omega, \tau)) f(a(\Omega, \tau), \varphi(\Omega, \tau)) d\varphi(\Omega, \tau) \end{aligned} \quad (5.82)$$

mit der selben Überlegung, wie in der Gleichung 5.64, wobei  $\varphi(\Omega, \tau)$  den Phasenverlauf darstellt. Damit ergibt sich der MMSE-Schätzer

$$\begin{aligned} & \hat{A}(\Omega, \tau) \\ &= \frac{\int_0^{\infty} \int_0^{2\pi} a(\Omega, \tau) f(Y(\Omega, \tau) | a(\Omega, \tau), \varphi(\Omega, \tau)) f(a(\Omega, \tau), \varphi(\Omega, \tau)) d\varphi(\Omega, \tau) da(\Omega, \tau)}{\int_0^{\infty} \int_0^{2\pi} f(Y(\Omega, \tau) | a(\Omega, \tau), \varphi(\Omega, \tau)) f(a(\Omega, \tau), \varphi(\Omega, \tau)) d\varphi(\Omega, \tau) da(\Omega, \tau)}. \end{aligned} \quad (5.83)$$

Nach dem Einsetzen von  $f(Y(\Omega, \tau) | A(\Omega, \tau), \varphi(\Omega, \tau))$  wie in der Gleichung 5.62 und

$$f(A(\Omega, \tau), \varphi(\Omega, \tau)) = \frac{A(\Omega, \tau)}{2\pi\sigma_X^2} \exp\left\{-\frac{A(\Omega, \tau)^2}{4\sigma_X^2}\right\} \quad (5.84)$$

$$\sigma_X^2 = E[|X(\Omega, \tau)|^2] \quad (5.85)$$

wie in der Gleichung 4.28 erhält man einen geschlossenen Ausdruck für den Schätzer

$$\begin{aligned} \hat{A}(\Omega, \tau) &= \Gamma(1,5) \frac{\sqrt{\nu(\Omega, \tau)}}{\gamma(\Omega, \tau)} \exp\left(-\frac{\nu(\Omega, \tau)}{2}\right) \\ &\times \left[ (1 + \nu(\Omega, \tau)) I_0\left(\frac{\nu(\Omega, \tau)}{2}\right) + \nu(\Omega, \tau) I_1\left(\frac{\nu(\Omega, \tau)}{2}\right) \right] R(\Omega, \tau) \end{aligned} \quad (5.86)$$

$\Gamma(\cdot)$  bezeichnet dabei die Gamma-Funktion,  $I_0(\cdot)$  und  $I_1(\cdot)$  stehen für die modifizierten Besselfunktionen nullter und erster Ordnung,

$$\nu(\Omega, \tau) = (1 + \gamma(\Omega, \tau)) \frac{\xi(\Omega, \tau)}{1 + \xi(\Omega, \tau)} \quad (5.87)$$

und  $\gamma(\Omega, \tau)$  und  $\xi(\Omega, \tau)$  sind die a-posteriori und A-Priori-SNR, wie in den Gleichungen 5.56 und 5.49.

Eine Alternative zur MMSE-Schätzung wurde in [EM85] gegeben. Dabei wird statt der Differenz der Beträge, wie in der Gleichung 5.76, die Differenz der logarithmierten Amplituden minimiert

$$E \left[ \left( \log A(\Omega, \tau) - \log \hat{A}(\Omega, \tau) \right)^2 \right] \stackrel{!}{=} \min. \quad (5.88)$$

Die daraus resultierende Schätzung soll die logarithmische Leistungswahrnehmung des menschlichen Gehörs stärker berücksichtigen.

Mit den oben eingeführten Dichtemodellen ergibt sich die MMSE-LSA-Schätzung für die logarithmierte Amplitude

$$\hat{A}(\Omega, \tau) = \frac{\xi(\Omega, \tau)}{\xi(\Omega, \tau) + 1} \exp\left(\frac{1}{2} \int_{\nu(\Omega, \tau)}^{\infty} \frac{e^{-t}}{t} dt\right) R(\Omega, \tau). \quad (5.89)$$

Die MMSE-LSA-Lösung weicht nur geringfügig von der MMSE-Lösung ab. Viel bedeutender wirken sich die Schätzungen der Rauschleistung und der A-Priori-SNR auf die Qualität der Spektralen Subtraktion aus.

Die einzelnen Übertragungsfunktionen der hier besprochenen Algorithmen zur Spektralen Subtraktion sind noch ein mal in der Tabelle 5.1 zusammengefasst.

**Tabelle 5.1.:** Übertragungsfunktionen der einzelnen Algorithmen zur Spektralen Subtraktion.

Algorithmus	Übertragungsfunktion
Wiener-Filter	$G^W(\Omega, \tau) = \frac{\xi(\Omega, \tau)}{\xi(\Omega, \tau) + 1}$
Parametrisches Wiener-Filter	$G^{pW}(\Omega, \tau) = \left( \frac{\xi(\Omega, \tau)}{\xi(\Omega, \tau) + a} \right)^\beta$
Maximum-Likelihood-Subtraktion	$G^{ML}(\Omega, \tau) = \frac{1}{2} + \frac{1}{2} \sqrt{\frac{\xi(\Omega, \tau)}{\xi(\Omega, \tau) + 1}}$
MMSE	$G^{MMSE}(\Omega, \tau) = \Gamma(1, 5) \frac{\sqrt{\nu(\Omega, \tau)}}{\gamma(\Omega, \tau)} \exp\left(-\frac{\nu(\Omega, \tau)}{2}\right) \times \left[ (1 + \nu(\Omega, \tau)) I_0\left(\frac{\nu(\Omega, \tau)}{2}\right) + \nu(\Omega, \tau) I_1\left(\frac{\nu(\Omega, \tau)}{2}\right) \right]$
MMSE-LSA	$G^{MMSE-LSA}(\Omega, \tau) = \frac{\xi(\Omega, \tau)}{\xi(\Omega, \tau) + 1} \exp\left(\frac{1}{2} \int_{\nu(\Omega, \tau)}^{\infty} \frac{e^{-t}}{t} dt\right)$

### 5.3.3. Schätzung des A-Priori SNR

Wie der Tabelle 5.1 zu entnehmen ist, basieren die Berechnungen der Übertragungsfunktionen auf den A-Priori und A-Posteriori SNR-Werten. Während das A-Posteriori SNR (Gl. 5.56) eine lokale Schätzung des aktuellen SNR ist, wird das A-Priori SNR durch Erwartungswertbildung über die Amplituden des reinen Signals gegeben. Da zu jedem Zeitpunkt nur das verrauschte Signal vorliegt, müssen das A-Priori SNR und das a-posteriori SNR geschätzt werden.

Wie in [EM84] gezeigt wurde, ist die Übertragungsfunktion des MMSE-Schätzers sehr sensitiv gegenüber kleinen Schwankungen des A-Priori SNR. Im Laufe der letzten Jahre wurden mehrere Methoden für die Schätzung von A-Priori SNR vorgeschlagen [EM84, SK00, HSK04, Coh05]. Hier wird die wahrscheinlich bekannteste diskutiert.

Unter der Annahme, dass das Sprach- und das Rauschsignal unabhängige Gaussvariablen sind (vgl. Abschnitt 5.3.2.1), ist der Zusammenhang zwischen A-Priori und a-posteriori SNR gegeben durch

$$\xi(\Omega, \tau) = \frac{E(A^2(\Omega, \tau))}{\sigma_D^2(\Omega, \tau)} \quad (5.90)$$

$$= \frac{E(R^2(\Omega, \tau) - |D(\Omega, \tau)|^2)}{\sigma_D^2(\Omega, \tau)} \quad (5.91)$$

$$= \frac{E(R^2(\Omega, \tau))}{\sigma_D^2(\Omega, \tau)} - \frac{E(|D(\Omega, \tau)|^2)}{\sigma_D^2(\Omega, \tau)} \quad (5.92)$$

$$= E(\gamma(\Omega, \tau)) - 1 \quad (5.93)$$

wobei  $A(\Omega, \tau)$ ,  $R(\Omega, \tau)$  und  $D(\Omega, \tau)$  die Amplitudenverläufe des Nutz-, Mikrofon-

und Rauschsignals bezeichnen und  $\sigma_D^2(\Omega, \tau)$  die Varianz und  $\gamma(\Omega, \tau)$  das a-posteriori SNR darstellen. Kombiniert man die Gleichung 5.93 mit 5.49, so kann das A-Priori SNR aus dem a-posteriori SNR und der spektralen Amplitude stochastisch berechnet werden

$$\xi(\Omega, \tau) = \alpha \frac{\hat{A}^2(\Omega, \tau - 1)}{\sigma_D^2(\Omega, \tau - 1)} + (1 - \alpha) \max[\gamma(\Omega, \tau) - 1, 0] \quad (5.94)$$

$$= \alpha G^2(\Omega, \tau - 1) \frac{|Y^2(\Omega, \tau - 1)|}{\sigma_D^2(\Omega, \tau - 1)} + (1 - \alpha) \max[\gamma(\Omega, \tau) - 1, 0] \quad (5.95)$$

mit  $0 < \alpha < 1$ . Der Initialwert von  $\xi(\Omega, 0)$  wird gegeben durch

$$\xi(\Omega, 0) = \alpha + (1 - \alpha) \max[\gamma(\Omega, 0) - 1, 0]. \quad (5.96)$$

Bei diesem Ansatz wird die Schätzung des A-Priori SNR an den Stellen des inaktiven Sprachsignals zu einer stark geglätteten Schätzung des a-posteriori SNRs. Hingegen an den Stellen der leistungsstarken Sprachaktivität folgt die Schätzung des A-Priori SNRs fast ungeglättet dem a-posteriori SNR. Damit können im Idealfall einerseits die Musical Tones in den Sprachpausen vermieden werden, andererseits bleibt das Sprachsignal unverhallt und unverzerrt.

### 5.3.4. Wahrscheinlichkeit für Sprachaktivität

Bei der Herleitung der oben vorgestellten Übertragungsfunktionen wurde stets davon ausgegangen, dass das Sprachsignal permanent aktiv ist. In der Realität ist das nicht immer der Fall. Sprachpausen können selbst in den Signalabschnitten auftreten, in denen das Sprachsignal aktiv ist. Z.B. während ein Vokal gesprochen wird, ist die Signalaktivität nur in den Frequenzbändern zu beobachten, die die Vielfachen der Grundfrequenz bilden (Harmonischen).

Bei der Bestimmung der Übertragungsfunktion der Spektralen Subtraktion kann also zusätzlich die Wahrscheinlichkeit der Sprachaktivität berücksichtigt werden, was zu einer Verbesserung der Rauschsignalreduktion führt [MM80, Coh03]. Dazu wird von einem Zwei-Zustands-Modell ausgegangen, bei dem im Signal die Sprache entweder aktiv ist oder nicht. Mathematisch werden somit zwei Hypothesen angenommen<sup>5</sup>:

---

<sup>5</sup> $H_k(\Omega, \tau)$  ist die Bezeichnung für die Hypothese  $k$  und darf nicht mit der Entropie verwechselt werden. Die Bezeichnung  $H_k(\Omega, \tau)$  hat sich in der Fachliteratur durchgesetzt und wurde deswegen für die vorliegende Arbeit übernommen. Da aber die  $H_k(\Omega, \tau)$  und die Entropie in dieser Arbeit nicht in dem selben Kontext vorkommen, ist die Verwechslungsgefahr relativ klein.

- $H_0(\Omega, \tau)$ : Sprache ist inaktiv, d.h. die Gleichung 5.35 wird zu  $Y(\Omega, \tau) = D(\Omega, \tau)$  oder
- $H_1(\Omega, \tau)$ : Sprache ist aktiv, d.h.  $Y(\Omega, \tau) = X(\Omega, \tau) + D(\Omega, \tau)$ .

Liegt das Wissen über die Zustände  $H_0$  und  $H_1$  in Form von Wahrscheinlichkeiten für die Sprachaktivität bzw. Sprachabwesenheit vor, so wird der Schätzer für die Spektralampplituden des Sprachsignals aus der Gleichung 5.78 zu

$$\hat{A}(\Omega, \tau) = E[A(\Omega, \tau) | Y(\Omega, \tau)] \quad (5.97)$$

$$= E[A(\Omega, \tau) | Y(\Omega, \tau), H_1(\Omega, \tau)] P(H_1(\Omega, \tau) | Y(\Omega, \tau)) \\ + E[A(\Omega, \tau) | Y(\Omega, \tau), H_0(\Omega, \tau)] P(H_0(\Omega, \tau) | Y(\Omega, \tau)) \quad (5.98)$$

mit  $P(H_k(\Omega, \tau) | Y(\Omega, \tau))$  als Wahrscheinlichkeit, dass das Sprachsignal im Zustand  $H_k(\Omega, \tau)$  ist, gegeben das gemessene Signal  $Y(\Omega, \tau)$ . Der Term  $E[A(\Omega, \tau) | Y(\Omega, \tau), H_0(\Omega, \tau)]$  repräsentiert den Erwartungswert der Spektralampplituden  $A(\Omega, \tau)$  gegeben das gemessene Signal  $Y(\Omega, \tau)$  UND der Zustand  $H_0(\Omega, \tau)$  in dem die Sprache inaktiv ist, d.h. nur die Rauschkomponenten im Signal vorhanden sind. Diese gilt es zu minimieren. Dies geschieht, indem man für die Übertragungsfunktion  $G_{H_0}(\Omega, \tau)$  pauschal einen Wert  $G_{\min}$  angibt, der für die Unterdrückung der Rauschanteile in den Sprachpausen sorgt (z.B.  $G_{\min} = 0$ , [MM80]). Die Berechnung von  $E[A(\Omega, \tau) | Y(\Omega, \tau), H_1(\Omega, \tau)]$  läuft auf die Berechnung der Übertragungsfunktion des Filters in den Signalperioden hinaus, in denen die Sprache aktiv ist, so wie das in den oberen Abschnitten beschrieben wurde. So kann z.B. dafür die Gleichung 5.75 angesetzt werden

$$E[A(\Omega, \tau) | Y(\Omega, \tau), H_1(\Omega, \tau)] = \left( \frac{1}{2} + \frac{1}{2} \sqrt{\frac{\xi(\Omega, \tau)}{\xi(\Omega, \tau) + 1}} \right) R(\Omega, \tau) \quad (5.99)$$

$$= G_{H_1}(\Omega, \tau) R(\Omega, \tau). \quad (5.100)$$

Damit ergibt sich die Übertragungsfunktion  $G(\Omega, \tau)$

$$G(\Omega, \tau) = G_{H_1}(\Omega, \tau) P(H_1(\Omega, \tau) | Y(\Omega, \tau)) + G_{\min} P(H_0(\Omega, \tau) | Y(\Omega, \tau)) \quad (5.101)$$

bzw. mit  $P(H_0(\Omega, \tau) | Y(\Omega, \tau)) = 1 - P(H_1(\Omega, \tau) | Y(\Omega, \tau))$

$$G(\Omega, \tau) = G_{H_1}(\Omega, \tau) P(H_1(\Omega, \tau) | Y(\Omega, \tau)) \\ + G_{\min}(1 - P(H_1(\Omega, \tau) | Y(\Omega, \tau))). \quad (5.102)$$

Diese Form kann grundsätzlich für alle Methoden der Spektralen Subtraktion benutzt werden. Eine Ausnahme bildet dabei der MMSE-LSA-Algorithmus, da bei diesem von der Schätzung von  $\log \hat{A}(\Omega, \tau)$  ausgegangen wird [CB01]. Damit wird  $\hat{A}^{MMSE-LSA}(\Omega, \tau)$  zu

$$\begin{aligned} \hat{A}^{MMSE-LSA}(\Omega, \tau) &= \exp \{E [\log(A(\Omega, \tau)) | Y(\Omega, \tau), H_1(\Omega, \tau)] P(H_1(\Omega, \tau) | Y(\Omega, \tau)) \\ &\quad + E [\log(A(\Omega, \tau)) | Y(\Omega, \tau), H_0(\Omega, \tau)] P(H_0(\Omega, \tau) | Y(\Omega, \tau))\} \\ &= \exp \{E [\log(A(\Omega, \tau)) | Y(\Omega, \tau), H_1(\Omega, \tau)]\}^{P(H_1(\Omega, \tau) | Y(\Omega, \tau))} \\ &\quad \cdot \exp \{E [\log(A(\Omega, \tau)) | Y(\Omega, \tau), H_0(\Omega, \tau)]\}^{P(H_0(\Omega, \tau) | Y(\Omega, \tau))}. \end{aligned} \quad (5.103)$$

Daraus resultiert die Übertragungsfunktion vom MMSE-LSA-Algorithmus

$$G^{MMSE-LSA}(\Omega, \tau) = G_{H_1}(\Omega, \tau)^{P(H_1(\Omega, \tau) | Y(\Omega, \tau))} G_{\min}^{(1-P(H_1(\Omega, \tau) | Y(\Omega, \tau)))}. \quad (5.104)$$

Die Berechnung der Wahrscheinlichkeit für Sprachaktivität  $P(H_1(\Omega, \tau) | Y(\Omega, \tau))$  ist dabei ein offenes Problem und wird in den folgenden Abschnitten behandelt. Die Wahrscheinlichkeiten werden aus den Parametern des verrauschten Signals, wie z.B. A-Priori- und a-posteriori-SNR, gewonnen, die dann in Abhängigkeit von den errechneten Wahrscheinlichkeiten in der nächsten Iteration aktualisiert werden können (Abbildung 5.9). Somit geschieht die Reduktion der Rauschanteile im Signal in zwei Schritten. Dieser Vorgang ist mit einer erhöhten Rechenzeit verbunden, die man aber in Kauf nimmt, um eine bessere Signalqualität zu gewinnen.

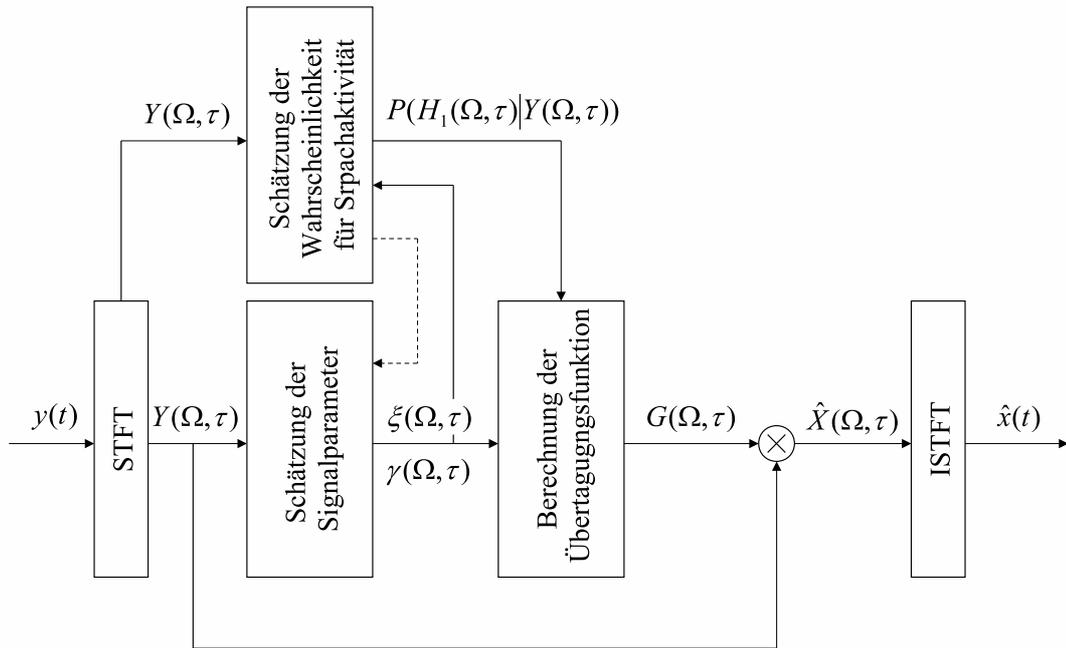
Die Unterteilung des Signals in die Perioden mit aktiven Sprachanteilen und in die Perioden ohne Sprache ermöglicht weiterhin eine genauere Schätzung der Rauschleistung. Unter der Annahme, dass der Rauschanteil über längere Zeit stationär bleibt, lässt sich eine Fallunterscheidung

$$H_0(\Omega, \tau) : \sigma_D^2(\Omega, \tau) = \alpha_D \sigma_D^2(\Omega, \tau - 1) + (1 - \alpha_D) |Y(\Omega, \tau)|^2 \quad (5.105)$$

$$H_1(\Omega, \tau) : \sigma_D^2(\Omega, \tau) = \sigma_D^2(\Omega, \tau - 1) \quad (5.106)$$

für die Schätzung der Rauschleistung einführen [Coh03]. Hierbei bezeichnet  $\alpha_D$  ( $0 < \alpha_D < 1$ ) einen Glättungsfaktor. Da die Information über die Hypothesen  $H_0$  und  $H_1$  in Form von bedingten Wahrscheinlichkeiten vorliegt, lassen sich letztere dazu verwenden, eine rekursive Form für die Berechnung der Rauschleistung zu definieren

$$\begin{aligned} \sigma_D^2(\Omega, \tau) &= P(H_1(\Omega, \tau) | Y(\Omega, \tau)) \sigma_D^2(\Omega, \tau - 1) \\ &\quad + (1 - P(H_1(\Omega, \tau) | Y(\Omega, \tau))) [\alpha_D \sigma_D^2(\Omega, \tau - 1) + (1 - \alpha_D) |Y(\Omega, \tau)|^2] \end{aligned} \quad (5.107)$$



**Abbildung 5.9.:** Datenfluss bei der Spektralen Subtraktion unter Berücksichtigung der Wahrscheinlichkeit für Sprachaktivität

bzw.

$$\sigma_D^2(\Omega, \tau) = \tilde{\alpha}_D \sigma_D^2(\Omega, \tau - 1) + (1 - \tilde{\alpha}_D) |Y(\Omega, \tau)|^2 \quad (5.108)$$

mit

$$\tilde{\alpha}_D = \alpha_D + (1 - \alpha_D) P(H_1(\Omega, \tau) | Y(\Omega, \tau)). \quad (5.109)$$

Im Folgenden werden die zwei bekanntesten Methoden zur Berechnung der Sprachwahrscheinlichkeit vorgestellt.

#### 5.3.4.1. Berechnung der Sprachwahrscheinlichkeit nach McAulay-Mallpass

Einer der ersten Ansätze für die Berechnung der Sprachwahrscheinlichkeiten wurde von McAulay-Mallpass gegeben [MM80]. Dabei geht man von dem Bayestheorem

für die Wahrscheinlichkeiten aus

$$P(H_1(\Omega, \tau) | Y(\Omega, \tau)) = P(H_1(\Omega, \tau) | R(\Omega, \tau)) = \frac{f(R(\Omega, \tau) | H_1(\Omega, \tau))P(H_1(\Omega, \tau))}{f(R(\Omega, \tau) | H_1(\Omega, \tau))P(H_1(\Omega, \tau)) + f(R(\Omega, \tau) | H_0(\Omega, \tau))P(H_0(\Omega, \tau))}. \quad (5.110)$$

Dabei bezeichnet  $f(R(\Omega, \tau) | H_k(\Omega, \tau))$  die *a priori* Verteilungsdichte des gemessenen Signals, gegeben die Hypothese  $H_k(\Omega, \tau)$  und  $P(H_k(\Omega, \tau))$  die Wahrscheinlichkeit des Zustandes  $H_k(\Omega, \tau)$ . Da diese nicht bekannt sind, geht man davon aus, dass sowohl die Sprache als auch das Rauschen gleich wahrscheinlich

$$P(H_1(\Omega, \tau)) = P(H_0(\Omega, \tau)) = \frac{1}{2} \quad (5.111)$$

sind. Für die Hypothese  $H_0$  ( $Y(\Omega, \tau) = D(\Omega, \tau)$ , bzw.  $R(\Omega, \tau) = |D(\Omega, \tau)|$ ) ergibt sich unter der Annahme der unabhängigen gaußverteilten Real- und Imaginärteile des Rauschens die Rayleigh-Verteilung für die Amplituden (mit der selben Motivation wie in den Abschnitten 4.3.1 und 5.3.2.3)

$$f(R(\Omega, \tau) | H_0(\Omega, \tau)) = \frac{2R(\Omega, \tau)}{\pi\sigma_D^2} \exp\left\{-\frac{R(\Omega, \tau)^2}{\sigma_D^2}\right\}. \quad (5.112)$$

Für die Hypothese  $H_1$  ( $Y(\Omega, \tau) = X(\Omega, \tau) + D(\Omega, \tau)$ ) resultiert aus der Gleichung 5.65 unter der Berücksichtigung der Transformationsvorschrift 4.25 die Rice-Verteilung der Amplituden (Abschnitt 2.1.4.6)

$$f(R(\Omega, \tau) | H_1(\Omega, \tau)) = \frac{2R(\Omega, \tau)}{\pi\sigma_D^2} \exp\left(-\frac{R^2(\Omega, \tau) + A^2(\Omega, \tau)}{\sigma_D^2}\right) I_0\left(\frac{2R(\Omega, \tau)A(\Omega, \tau)}{\sigma_D^2}\right). \quad (5.113)$$

Setzt man nun die Gleichungen 5.111-5.113 in die Gleichung 5.110 ein, ergibt sich die Wahrscheinlichkeit der Sprachaktivität

$$P(H_1(\Omega, \tau) | R(\Omega, \tau)) = \frac{\exp(-\xi(\Omega, \tau)) I_0\left(2\sqrt{\xi(\Omega, \tau)\frac{R^2(\Omega, \tau)}{\sigma_D^2}}\right)}{1 + \exp(-\xi(\Omega, \tau)) I_0\left(2\sqrt{\xi(\Omega, \tau)\frac{R^2(\Omega, \tau)}{\sigma_D^2}}\right)}. \quad (5.114)$$

### 5.3.4.2. Berechnung der Sprachwahrscheinlichkeit nach I. Cohen

Alternativ kann man die Sprachwahrscheinlichkeit aus den bedingten Verteilungen der geschätzten SNRs gewinnen [Coh03]. Seien

$$f(Y(\Omega, \tau) | H_0(\Omega, \tau)) = \frac{1}{\pi \sigma_D^2} \exp \left\{ -\frac{|Y(\Omega, \tau)|^2}{\sigma_D^2} \right\} \quad (5.115)$$

und

$$f(Y(\Omega, \tau) | H_1(\Omega, \tau)) = \frac{1}{\pi(\sigma_D^2 + \sigma_X^2)} \exp \left\{ -\frac{|Y(\Omega, \tau)|^2}{\sigma_D^2 + \sigma_X^2} \right\}, \quad (5.116)$$

die bedingten Verteilungsdichten des gemessenen Signals, gegeben die Hypothese  $H_k(\Omega, \tau)$ , wobei  $\sigma_X^2$  die Varianz des Nutzsignals ist. Dann können die entsprechenden Verteilungsdichten des a posteriori SNR  $\gamma(\Omega, \tau)$  nach Gl. 5.56 unter der Berücksichtigung der Koordinatentransformation als

$$f(\gamma(\Omega, \tau) | H_0(\Omega, \tau)) = \begin{cases} 0 & x < 0 \\ \exp \{-\gamma(\Omega, \tau)\} & \text{sonst} \end{cases} \quad (5.117)$$

und

$$f(\gamma(\Omega, \tau) | H_1(\Omega, \tau)) = \begin{cases} 0 & x < 0 \\ \frac{1}{1+\xi(\Omega, \tau)} \exp \left\{ -\frac{\gamma(\Omega, \tau)}{1+\xi(\Omega, \tau)} \right\} & \text{sonst} \end{cases} \quad (5.118)$$

definiert werden. Wie man sieht folgt das a posteriori SNR der Laplace-Verteilung (Abschnitt 2.1.4.2). Wendet man nun das Bayestheorem (Gl. 5.110) für die bedingte Wahrscheinlichkeit  $P(H_1(\Omega, \tau) | \gamma(\Omega, \tau))$  an, so erhält man

$$P(H_1(\Omega, \tau) | \gamma(\Omega, \tau)) = \frac{1}{1 + \frac{P(H_0(\Omega, \tau))}{P(H_1(\Omega, \tau))} (1 + \xi(\Omega, \tau)) \exp(-\nu(\Omega, \tau))} \quad (5.119)$$

mit  $\nu(\Omega, \tau)$  wie in der Gleichung 5.87.

## 5.4. Anwendung der Methoden der Spektralen Subtraktion im mehrkanaligen Szenario

Die Verfahren der einkanaligen Spektralen Subtraktion lassen sich ohne weiteres auf die unabhängigen Komponenten bzw. auf die durch ICA oder Beamformer getrennten Kanäle einzeln anwenden. Dabei werden, wie in einem einkanaligen Fall, die

Rauschanteile in den Signalen minimiert. Um die Qualität der Rauschreduktion zu verbessern, können die Informationen aus der höheren Anzahl von Kanälen in Betracht gezogen werden. Vor allem können diese bei der Schätzung der Rauschleistung und den damit verbundenen A-Priori- und a-posteriori-SNR und bei der Berechnung der Wahrscheinlichkeiten für Sprachaktivität ausgenutzt werden. Die Bestimmung der Übertragungsfunktion wird nur indirekt von den mehrkanaligen Informationen beeinflusst, da man hier von dem selben Sprachmodell ausgeht, wie im einkanaligen Fall.

An dieser Stelle wird vom Signalmodell ausgegangen, das in der Gleichung 4.4 formuliert wurde. Die Lösung des Problems bilden die statistisch unabhängigen Komponenten  $\mathbf{Y}(\Omega, \tau)$ , für die in jedem Frequenzband das Permutationsproblem gelöst wurde und die bleibende Interferenzen von anderen Kanälen enthalten, die es zu minimieren gilt. Somit genügt jedes einzelne Signal des Vektors dem additiven Modell aus der Gleichung 5.35, in dem das Nutzsinal  $Y_i(\Omega, \tau)$  im  $i$ -ten Kanal durch ein additives Rauschsignal gestört wird. Dadurch können die einzelnen Kanäle als gegenseitige Störsignale angesehen werden. Geht man davon aus, dass die Trennung gut funktioniert hat, d.h. dass die Interferenzanteile im jeweiligen Signal kleiner sind als die Nutzanteile und dass die Einfallsrichtung jedes Signals eindeutig aus der geschätzten Mischungsmatrix bestimmbar ist, so können diese bei der Berechnung der Sprachwahrscheinlichkeiten ausgenutzt werden.

### 5.4.1. Berechnung der Sprachwahrscheinlichkeit

Wie oben schon erklärt wurde, versteht man unter der Wahrscheinlichkeit für die Sprachaktivität in einem einkanaligen Fall die Wahrscheinlichkeit  $P(H_1(\Omega, \tau) | Y(\Omega, \tau))$  für die Hypothese  $H_1(\Omega, \tau)$ , gegeben die Messdaten. Solche Wahrscheinlichkeiten gilt es unter der Berücksichtigung der mehrkanaligen Informationen für jede unabhängige Komponente  $Y_i(\Omega, \tau)$  zu bestimmen. Ein ähnliches Ziel stellt man sich bei der Bestimmung der Maskenfunktionen aus den Abschnitten 5.1-5.2. Vergleicht man die Eigenschaften der Zeit-Frequenz-Masken mit den Eigenschaften von  $P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau))$ , so stellt man fest, dass die Maskenfunktionen und die Wahrscheinlichkeit für die Sprachaktivität korrespondieren:

- der Wertebereich sowohl von den Zeit-Frequenz-Masken, als auch von  $P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau))$  liegt im Intervall  $[0,1]$ ,
- sowohl die Zeit-Frequenz-Masken, als auch  $P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau))$  gehen gegen 1, wenn das Nutzsinal (sprich  $S_i(\Omega, \tau)$ ) aktiv bzw. wenn das Nutzsinal

dominant ist, und gehen gegen 0, an den Stellen im Signal, wo nur Rauschen vorhanden ist bzw. wo keine Aktivität von  $S_i(\Omega, \tau)$  zu detektieren ist.

Die Idee für die Bestimmung der Wahrscheinlichkeiten ist nun, die Maskierungsfunktionen als die Wahrscheinlichkeit für die Sprachaktivität, gegeben die Daten, zu interpretieren. Damit ergeben sich mehrere mögliche Varianten für die Berechnung von  $P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau))$ . Um die Auswahl etwas einzuschränken, werden im Weiteren folgende Masken betrachtet

- $P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau)) = M_i^{\text{DUE}}(\Omega, \tau)$  aus der Gleichung 5.10,
- $P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau)) = M_i^{\text{Amp}}(\Omega, \tau)$  aus der Gleichung 5.13,
- $P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau)) = M_i^{\text{Phase}}(\Omega, \tau)$  aus der Gleichung 5.20 und
- $P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau)) = M_i^{\text{Int}}(\Omega, \tau)$  aus der Gleichung 5.27.

$P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau))$  bildet ein Maß für die Dominanz der Quelle  $i$  in der Mischung. So wäre es möglich, für die Bestimmung von  $P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau))$  das Leistungsverhältnis aus der Gleichung 4.8 zu benutzen, da es genau die oben genannten Eigenschaften von  $P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau))$  hat. Damit wäre eine weitere Möglichkeit für die Bestimmung von Sprachwahrscheinlichkeiten durch  $P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau)) = \text{powerRatio}_i(\Omega, \tau)$  gegeben.

### 5.4.2. Mehrkanalige Störgeräuschunterdrückung unter Berücksichtigung der Sprachwahrscheinlichkeiten

Zusammenfassend wird im Folgenden ein Algorithmus zur mehrkanaligen Störgeräuschunterdrückung unter Berücksichtigung der Sprachwahrscheinlichkeiten gegeben. Wichtige Punkte neben der Berechnung der Sprachwahrscheinlichkeiten sind dabei die Schätzung der Rauschleistung  $(\sigma_D^{(i)}(\Omega, \tau))^2$  im  $i$ -ten Kanal und die Schätzung der a priori SNR  $\xi_i(\Omega, \tau)$  für die  $i$ -te unabhängige Komponente  $Y_i(\Omega, \tau - 1)$ . Diese müssen an die mehrkanalige Situation angepasst werden, damit nehmen die Gleichungen 5.107 und 5.95 die Form

$$\begin{aligned} (\sigma_D^{(i)}(\Omega, \tau))^2 &= P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau)) (\sigma_D^{(i)}(\Omega, \tau - 1))^2 \\ &\quad + (1 - P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau))) \\ &\quad \times \left[ \alpha_D (\sigma_D^{(i)}(\Omega, \tau - 1))^2 + (1 - \alpha_D) |Y_i(\Omega, \tau)|^2 \right] \end{aligned} \quad (5.120)$$

und

$$\xi_i(\Omega, \tau) = \alpha G_i^2(\Omega, \tau - 1) \frac{|Y_i^2(\Omega, \tau - 1)|}{\left(\sigma_D^{(i)}(\Omega, \tau - 1)\right)^2} + (1 - \alpha) \max[\gamma_i(\Omega, \tau) - 1, 0] \quad (5.121)$$

an. Die Letztere kann alternativ durch die neu gewonnenen Wahrscheinlichkeiten für die Sprachaktivität umformuliert werden

$$\xi_i(\Omega, \tau) = \alpha \xi_i(\Omega, \tau - 1) + (1 - \alpha) \frac{|Y_i^2(\Omega, \tau)| P\left(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau)\right)}{\left(\sigma_D^{(i)}(\Omega, \tau - 1)\right)^2} \quad (5.122)$$

so, wie das in [HKO09] vorgeschlagen wurde. Damit ergibt sich der **Algorithmus 5**.

---

**Algorithmus 5:** Algorithmus zur Nachbearbeitung der getrennten Sprachsignale unter Berücksichtigung der Sprachwahrscheinlichkeiten.

---

```

begin
  Initialisierung.
  1   for 0 < k ≤ NFFT/2 + 1 do
  2     for 0 < i ≤ N do
  3       γi(Ωk, 0) = 1
  4       GH1(i)(Ωk, 0) = 1
  5       ξi(Ωk, 0) = α
  6       σD(i)(Ωk, 0) = |Yi(Ωk, 0)|2
     end
  end
  7   Berechne die Wahrscheinlichkeiten für die Sprachaktivität
  P(H1(i)(Ω, τ) | Y(Ω, τ)), so wie es im Abschnitt 5.4.1 beschrieben wurde (Gl.
  5.10, 5.13, 5.20, 5.27).
  8   for 0 < i ≤ N do
  9     for 0 < k ≤ NFFT/2 + 1 do
 10      for 0 < τ ≤ τEND do
 11        Berechne σD(i)(Ωk, τ) wie in der Gleichung 5.120.
 12        γi(Ωk, τ) =  $\frac{|Y_i(\Omega_k, \tau)|^2}{\sigma_D^{(i)}(\Omega_k, \tau)}$ ;
 13        Berechne ξi(Ωk, τ) wie in der Gleichung 5.121 bzw. in der Gleichung
 5.122.
 14        Berechne GH1(i)(Ωk, τ) wie in der Tabelle 5.1.
 15        Berechne G(i)(Ωk, τ) wie in der Gleichung 5.101 bzw. wie in der
  Gleichung 5.104, wenn der MMSE-LSA-Algorithmus benutzt wird.
      end
    end
  end
end

```

---



## 6. Bewertungsmethoden

Das Ziel der in der vorliegenden Arbeit vorgestellten Methoden der Permutationskorrektur sowie der Verfahren der Zeit-Frequenz-Maskierung ist im Endeffekt die Verbesserung der Signalqualität nach der blinden Quellentrennung. Dabei stellt sich die Frage nach den Bewertungskriterien und den entsprechenden Qualitätsmaßen. Diese werden in diesem Kapitel diskutiert bzw. neu definiert.

### 6.1. Bewertungsmethoden im Vergleich

Eine berechtigte Frage, die sich bei der Wahl der Algorithmen zur Verbesserung der Signalqualität stellt, ist die nach der weiteren Anwendung des Signals. Diese stellt bestimmte Anforderungen an die Ergebnisse der verwendeten Methoden und erfordert unter Umständen einen Vergleich unterschiedlicher Algorithmen miteinander. Ausgehend von den Wahlkriterien lassen sich unterschiedliche Qualitätsmaße definieren.

Im Bereich der Sprachsignalverarbeitung kann man grundsätzlich zwei Klassen von Bewertungsmethoden unterscheiden:

- die *objektiven* Bewertungsmethoden stellen an Hand eines objektiven Kriteriums (wie z.B. die Leistung der bleibenden Interferenzen bzw. des Rauschsignals) ein Maß für die Güte des Verfahrens dar. Diese Bewertungsmethoden ermöglichen somit einen direkten Vergleich zwischen unterschiedlichen Algorithmen und
- die *subjektiven* Bewertungsmethoden bieten eine Einschätzung der Qualität an Hand von in den Hörtests geprüften Signaleigenschaften, die von einer Gruppe von Probanden bewertet werden. Als Signaleigenschaften dienen üblicherweise erreichte Verständlichkeit, Signalverzerrung usw.

Da die subjektive Einschätzung der Qualität von vielen Kriterien abhängt und bei unterschiedlichen Personen unterschiedlich ist, erfordern die subjektiven Bewertungsmethoden eine breite Testpopulation und eine hohe Anzahl an durchgeführten Hörtests, um aussagekräftige Ergebnisse zu erreichen. Somit sind die Validierung und

der Vergleich der unterschiedlichen Verfahren an Hand von subjektiven Bewertungsmethoden mit einem sehr hohen Zeit- und Rechenaufwand verbunden. Aus diesem Grund wird die Qualität in den meisten Arbeiten durch die objektiven Bewertungsmethoden angegeben. Das Problem dabei ist, dass die meist verbreiteten objektiven Bewertungsmethoden, wie z.B. Signal-Interferenz-Verhältnis (s. Abschnitt 6.2.1) oder Signal-Verzerrung-Verhältnis (s. Abschnitt 6.2.1), wenig bzw. nichts über die erreichte Verständlichkeit und die wahrgenommenen Artefakte im Signal aussagen. Besonders auf dem Gebiet der Sprachsignalverarbeitung hängt der subjektive Höreindruck von zahlreichen Aspekten ab und stellt damit eine mehrdimensionale Größe dar, die nicht einfach mit den objektiven Bewertungsmethoden zu erfassen ist [VHH98]. Seit einiger Zeit gibt es Untersuchungen zur automatischen Signalauswertung ohne aufwendige Hörtests, die aber weitgehend die gleichen Qualitätsaussagen liefern [FSPZ07].

## 6.2. Die objektiven Bewertungsmethoden

Bei der objektiven Bewertung der Algorithmen zur blinden Quellentrennung stehen zwei Fragen im Vordergrund, nämlich:

- wie gut die Signalgemische getrennt wurden, und
- wie stark die geschätzten Quellsignale durch die Entmischung bzw. die Nachfilterung und Zeit-Frequenz-Maskierung der Signale verzerrt sind?

Da bei den realen Raumaufnahmen keine Informationen über die Mischungsart vorliegen, kann kein direkter Vergleich von Mischungs- und Entmischungsfiler durchgeführt werden. Aus diesem Grund werden für die Bewertungsmaße die Signalleistungen der einzelnen Quellen sowie der gemischten Signale benutzt. Dazu ist es nötig, die Quellsignale einzeln aufzunehmen, sodass die Informationen darüber vorhanden sind, wie jede einzelne Quelle am Mikrofonarray ankommt. Da die Mischung nach der Gleichung 4.1 als eine additive Überlagerung der gefilterten Quellen anzusehen ist, kann diese durch die Addition der Einzelsignale erzeugt werden, sodass eine realitätsnahe Mischung ohne Verlust der zur Auswertung benötigten Informationen erreicht wird [Sch01].

### 6.2.1. Signal-zu-Interferenz-Verhältnis

Ausgehend davon, dass die Informationen über die einzelnen Sprecher an jedem Mikrophon/Kanal vorliegen, ist es möglich Aussagen über die Dämpfung der Übersprecher bzw. der Interferenzen zu treffen. Ein Maß dafür ist das Signal-Interferenz-Verhältnis<sup>1</sup>. Das SIR wird gebildet als Verhältnis der Leistung des ungedämpften Ausgangssignals zur Summe der Leistungen der verbliebenen Interferenzen am selben Kanal

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

Dabei ist  $y_{i,s_i}(t)$  das  $i$ -te Ausgangssignal des Mischungs-Entmischungssystems, wenn nur die Quelle  $s_i$  aktiv ist, und  $E_t[\cdot]$  stellt den Erwartungswert über die Zeit dar [STS99]. Bei der Bewertung der Algorithmen der blinden Quellentrennung ist man jedoch nicht an den absoluten Werten des SIRs interessiert, sondern viel mehr an der SIR-Verbesserung

$$\Delta \text{SIR}_i = 10 \log_{10} \frac{E_t [y_{i,s_i}^2(t)]}{E_t \left[ \sum_{j \neq i} y_{i,s_j}^2(t) \right]} - 10 \log_{10} \frac{E_t [x_{i,s_i}^2(t)]}{E_t \left[ \sum_{j \neq i} x_{i,s_j}^2(t) \right]}. \quad (6.2)$$

Hierbei stellt der zweite Term der obigen Gleichung das Eingangs-SIR des gemischten Signals dar.

Es ist anzumerken, dass die Definitionen des SIR und  $\Delta \text{SIR}$  nur für die Umgebungen bzw. Aufnahmen gültig sind, bei denen kein Hintergrundrauschen vorhanden ist bzw. die Rauschkulisse vernachlässigbar klein ist. Bei den rauschbehafteten Aufnahmen ist stattdessen Signal-zu-Interferenz-und-Rauschen-Verhältnis<sup>2</sup> zu benutzen. Dieses benötigt jedoch zusätzliches Wissen über die Rauschsignale bzw. darüber, wie die Rauschsignale von einzelnen Mikrofonen aufgenommen werden. Das Ausgangs-SINR wird dann berechnet als

$$\text{SINR}_i^{\text{out}} = 10 \log_{10} \frac{E_t [y_{i,s_i}^2(t)]}{E_t \left[ \sum_{j \neq i} y_{i,s_j}^2(t) + \tilde{n}_{i,n}(t) \right]}. \quad (6.3)$$

Dabei ist  $\tilde{n}_{i,n}(t)$  Rauschsignal  $\mathbf{n}(t)$  gefiltert durch das Entmischungssystem am Kanal  $i$ , wenn keine andere Quelle aktiv ist. Analog zu  $\Delta \text{SIR}$  lässt sich das  $\Delta \text{SINR}$  angeben

<sup>1</sup>engl. *Signal-to-Interference-Ratio (SIR)*

<sup>2</sup>engl. *Signal-to-Interference-and-Noise-Ratio (SINR)*

als

$$\Delta\text{SINR}_i^{\text{out}} = 10 \log_{10} \frac{E_t [y_{i,s_i}^2(t)]}{E_t [\sum_{j \neq i} y_{i,s_j}^2(t) + \tilde{n}_{i,n}(t)]} - 10 \log_{10} \frac{E_t [x_{i,s_i}^2(t)]}{E_t [\sum_{j \neq i} x_{i,s_j}^2(t) + n_{i,n}(t)]}. \quad (6.4)$$

### 6.2.2. Signal-zu-Verzerrung-Verhältnis

Wie oben schon erwähnt wurde, stellt neben der Trennungsqualität die Verzerrung des Signals durch das Entmischungssystem ein wichtiges Bewertungskriterium dar. Diese ist jedoch nicht mit der Sprachqualität gleich zu setzen. Die Gründe dazu wurden bereits in Abschnitt 6.1 erläutert. Als ein Maß für die Verzerrung des Signals hat sich auf dem Gebiet der blinden Quellentrennung von Audiosignalen das Signal-Verzerrung-Verhältnis<sup>3</sup> durchgesetzt [AMSM05]. Dieses wird definiert als

$$\text{SDR}_i = 10 \log_{10} \frac{E_t [x_{i,s_i}^2(t)]}{E_t [(x_{i,s_i}(t) - \alpha y_{i,s_i}(t - D))^2]}, \quad (6.5)$$

mit  $y_{i,s_j}(t)$  und  $x_{i,s_j}(t)$  als Ausgangs- bzw. Mikrofonsignalen, wie sie im Abschnitt 6.2.1 definiert wurden, und  $\alpha$  und  $D$  als Konstanten, um die Amplitude und den Zeitversatz von  $y_{i,s_j}(t)$  auszugleichen, gewählt so, dass  $E_t [(x_{i,s_i}(t) - \alpha y_{i,s_i}(t - D))^2]$  minimiert wird.

Diese Definition von SDR korrespondiert mit der Definition der Signalverzerrung in [STS99].

Es sei anzumerken, dass die Verzerrung eines Signals unterschiedlich definiert werden kann. Es kann entweder wie oben als eine Abweichung des verarbeiteten Signals vom Original, oder als die Summe der Störkomponenten, wie Interferenzen, Rauschen und Artefakte verstanden werden. Das rechtfertigt eine alternative Definition des Signal-Verzerrung-Verhältnisses, die in [VJP06] gegeben wurde. Für die Bewertung der Verzerrung wird in der vorliegenden Arbeit das Signal-zu-Verzerrung-Verhältnis, wie in der Gleichung 6.5 verwendet.

---

<sup>3</sup>engl. *Signal-to-Distortion-Ratio (SDR)*

## 6.3. Die subjektiven Bewertungsmethoden

Wie oben schon beschrieben wurde, bieten die subjektiven Bewertungsmethoden eine Einschätzung der Qualität anhand von den in den Hörtests geprüften Signaleigenschaften. Die Durchführung der Hörtests und die Ermittlung der Signalqualität kann dabei durchaus unterschiedlich gestaltet werden. Die gewählte Signaleigenschaft (z.B. Verständlichkeit) kann z.B. dadurch beurteilt werden, dass man

- über die von mehreren Personen abgegebenen Wertungen mittelt (*Mean Opinion Score*),
- mehrere durch unterschiedliche Verfahren verarbeitete Signale miteinander vergleicht und bewertet (*Vergleichstest*),
- mit einem ungestörten Nutzsignal vergleicht und die Qualität einschätzt oder
- die Signaleigenschaft durch vorgegebene Attribute beschreibt (*Diagnostic Acceptability Measure*) [Pet02].

Möglich sind auch Kombinationen der oben vorgestellten Methoden zur subjektiven Bewertung.

Problematisch bei der Durchführung der Hörtests ist die Tatsache, dass je nach Sprecher-, Hörerauswahl und den Erwartungen der Hörerschaft die Ergebnisse bzw. die abgegebenen Bewertungen über die gleichen Hörbeispiele extrem streuen. Eine gewisse Vergleichbarkeit der Ergebnisse gelingt dadurch, dass man in die Tests Signale mit bekannten Bewertungen (Anker) integriert.

Auf diesem Prinzip basiert das Verfahren für die subjektive Bewertung, das in dieser Arbeit verwendet wurde. Es handelt sich dabei um das modifizierte *MUSHRA* (*Multiple Stimulus Test with Hidden Reference and Anchors*) [ITU03], das in [VJP06, JVA<sup>+</sup>08] speziell für die Bewertung von Algorithmen zur blinden Quellentrennung von Audiosignalen adaptiert wurde. Bei diesem Bewertungsverfahren werden einer Testperson mehrere Signale vorgespielt, unter denen sich ein Referenzsignal und mindestens ein Anker befinden. Die Testsignale sind mit einem Referenzsignal zu vergleichen und auf einer Skala von 0 bis 100 zu bewerten, wobei 100 immer dem Referenzsignal entspricht. Der Vergleich mit dem versteckten Referenzsignal ist unter anderem deswegen sinnvoll, weil man die unzuverlässigen Bewertungen, d.h. z.B. die Bewertungen, bei denen das Referenzsignal falsch eingeschätzt wurde, gleich aussortieren kann. Als Referenzsignal dient in diesem Fall ein Sprachsignal, das in der gegebenen Konstellation vom Mikrofonarray aufgenommen wird. Als Bewertungskriterien dienen:

- Sprachverständlichkeit,
- Unterdrückung der Störsignale und
- Verzerrung des Sprachsignals

bezogen auf das Referenzsignal.

# 7. Ergebnisse und Diskussion

In diesem Kapitel soll die Leistungsfähigkeit der vorgeschlagenen Methoden zur Permutationskorrektur sowie zur Zeit-Frequenz-Maskierung in Abhängigkeit von der jeweiligen akustischen Situation untersucht und mit Ergebnissen bereits existierender Verfahren verglichen werden. Die in dieser Hinsicht relevanten Parameter sind der Mikrofonabstand, die Nachhallzeit und die räumliche Quellenverteilung.

## 7.1. Quelldatensätze und Meßbedingungen

In der vorliegenden Arbeit wurden drei Sätze von Spachaufnahmen benutzt.

1. Der erste Satz wurde an der TU Berlin aufgenommen. Hier wurden die Daten aus der TIDIGITS-Datenbank [LDC93] in einem Büroraum in unterschiedlichen Konstellationen (Anzahl von Sprechern, Einfallswinkel und Abstand zum Mikrofonarray, Mikrofonabstand) aufgenommen. Die Abmessungen des Raumes betragen  $7,1 \times 6,7$  Meter bei einer Deckenhöhe von 3,5 Meter. Zum Zweck der Auswertung (siehe Abschnitt 6.2) wurden neben den Sprachmischungen auch die Einzelsignale der Sprecher aufgezeichnet. Weiterhin wurde für jede Aufnahme eine Reverberationszeit  $T_R$  gemessen<sup>1</sup>. Für die Aufnahmen wurden die Mikrofone des Typs Behringer ECM 8000 mit Tascam MA-8-Vorverstärker sowie ein Harddisk-Recorder Alesis HD24 verwendet. Die einzelnen Parameter der Mischungen sind in den Tabellen 7.1-7.3 zusammengefasst.
2. Der zweite Mischungssatz [Saw03] stammt ebenso aus einem Büroraum allerdings mit etwas kleineren Reverberationszeiten. Wie oben wurden hier unterschiedlichen Sprachsignale in Mischungen (bis zu vier Sprecher gleichzeitig) sowie einzeln aufgenommen. Die Raumdimensionen betragen  $3,55 \times 4,5$  Meter bei einer Deckenhöhe von 2,5 Meter, woraus sich eine Reverberationszeit

---

<sup>1</sup>Die Messungen von Reverberationszeiten basieren auf der Veröffentlichung von N. Aoshima [Aos81]

$T_R$  von 130 ms ergab [Saw03]. Die Parameter dieser Mischungen sind in der Tabelle 7.4 zu finden.

- Bei dem dritten Quelldatensatz wurden Sprachaufnahmen von vier Probanden gemacht, die sich in einer festen Position in Bezug auf das Mikrofonarray befanden und Sätze mit einer Länge von jeweils 20 Sekunden frei vorsprachen bzw. vorgelesen haben. Damit ergaben sich vier mehrkanalige Aufnahmen von Referenzsignalen, die dann zu Sprachmischungen von zwei bis vier Sprechern durch Superposition zusammengefügt wurden, so wie das von D. W. E. Schobben vorgeschlagen wurde<sup>2</sup> [Sch01]. Die Aufnahmen erfolgten in dem selben Raum und mit den selben Gerätschaften, wie im Punkt 1). Die einzelnen Parameter der Mischungen sind in der Tabelle 7.5 zusammengefasst.

Die Abbildung 7.1 zeigt den prinzipiellen Aufbau der Messungen.

**Tabelle 7.1.:** Mischungsparameter.

Mischung	Mix. 1 TU Berlin	Mix. 2 TU Berlin	Mix. 3 TU Berlin
Reverberationszeit $T_R$	159 ms	159 ms	159 ms
Mikrofonabstand $d$	3 cm	3 cm	3 cm
Abtastfrequenz $f_S$	11 kHz	11 kHz	11 kHz
Anzahl der Sprecher $N$	2	3	4
Distanz zwischen dem Sprecher $i$ und dem Arrayzentrum	$L_1 = L_2 = 0,9$ m	$L_1 = L_2 = L_3 = 0,9$ m	$L_1 = L_2 = L_3 = L_4 = 0,9$ m
Angulare Position vom Sprecher $i$ (wie in der Abbildung 7.1)	$\theta_1 = 50^\circ$ $\theta_2 = 115^\circ$	$\theta_1 = 30^\circ$ $\theta_2 = 80^\circ$ $\theta_3 = 135^\circ$	$\theta_1 = 25^\circ$ $\theta_2 = 80^\circ$ $\theta_3 = 130^\circ$ $\theta_4 = 155^\circ$
Eingang-SIR [dB]	-0,07	-3,05	-5,04

---

<sup>2</sup>Diese Vorgehensweise wurden genauer im Abschnitt 6.2 diskutiert

**Tabelle 7.2.:** Mischungsparameter.

Mischung	Mix. 4 TU Berlin	Mix. 5 TU Berlin	Mix. 6 TU Berlin
Reverberationszeit $T_R$	189 ms	189 ms	189 ms
Mikrofonabstand $d$	3 cm	3 cm	3 cm
Abtastfrequenz $f_S$	11 kHz	11 kHz	11 kHz
Anzahl der Sprecher $N$	2	3	4
Distanz zwischen dem Sprecher $i$ und dem Arrayzentrum	$L_1 = L_2 = 2,0$ m	$L_1 = L_2 = L_3 = 2,0$ m	$L_1 = L_2 = L_3 = L_4 = 2,0$ m
Angulare Position vom Sprecher $i$ (wie in der Abbildung 7.1)	$\theta_1 = 75^\circ$ $\theta_2 = 165^\circ$	$\theta_1 = 35^\circ$ $\theta_2 = 80^\circ$ $\theta_3 = 165^\circ$	$\theta_1 = 30^\circ$ $\theta_2 = 75^\circ$ $\theta_3 = 125^\circ$ $\theta_4 = 165^\circ$
Eingangs-SIR [dB]	-0,04	-3,4	-6,9

**Tabelle 7.3.:** Mischungsparameter.

Mischung	Mix. 7 TU Berlin	Mix. 8 TU Berlin	Mix. 9 TU Berlin
Reverberationszeit $T_R$	159 ms	159 ms	159 ms
Mikrofonabstand $d$	12 cm	12 cm	12 cm
Abtastfrequenz $f_S$	11 kHz	11 kHz	11 kHz
Anzahl der Sprecher $N$	2	3	4
Distanz zwischen dem Sprecher $i$ und dem Arrayzentrum	$L_1 = L_2 = 0,9$ m	$L_1 = L_2 = L_3 = 0,9$ m	$L_1 = L_2 = L_3 = L_4 = 0,9$ m
Angulare Position vom Sprecher $i$ (wie in der Abbildung 7.1)	$\theta_1 = 30^\circ$ $\theta_2 = 70^\circ$	$\theta_1 = 30^\circ$ $\theta_2 = 70^\circ$ $\theta_3 = 150^\circ$	$\theta_1 = 30^\circ$ $\theta_2 = 70^\circ$ $\theta_3 = 150^\circ$ $\theta_4 = 170^\circ$
Eingangs-SIR [dB]	0,02	-2,5	-4,2

## 7.2. Ergebnisse der Permutationskorrekturen

Die in dem Kapitel 4 vorgestellten Verfahren zur Permutationskorrektur wurden in Verbindung mit drei Algorithmen zur blinden Quellentrennung (JADE, FastICA und Infomax) an den Sprachmischungen aus den Tabellen 7.1-7.5 getestet. Dazu wurden die Audiosignale Kurzzeit-Fourier-transformiert mit der Auflösung von  $N_{FFT} = 1024$

**Tabelle 7.4.:** Mischungsparameter.

Mischung	Mix. 10 NTT[Saw03]	Mix. 11 NTT[Saw03]	Mix. 12 NTT[Saw03]
Reverberationszeit $T_R$	130 ms	130 ms	130 ms
Mikrofonabstand $d$	4 cm	4 cm	4 cm
Abtastfrequenz $f_S$	8 kHz	8 kHz	8 kHz
Anzahl der Sprecher $N$	2	3	4
Distanz zwischen dem Sprecher $i$ und dem Arrayzentrum	$L_1 = L_2 = 1,2$ m	$L_1 = L_2 = L_3 = 1,2$ m	$L_1 = L_2 = L_3 = L_4 = 1,2$ m
Angulare Position vom Sprecher $i$ (wie in der Abbildung 7.1)	$\theta_1 = 30^\circ$ $\theta_2 = 70^\circ$	$\theta_1 = 30^\circ$ $\theta_2 = 70^\circ$ $\theta_3 = 110^\circ$	$\theta_1 = 30^\circ$ $\theta_2 = 70^\circ$ $\theta_3 = 110^\circ$ $\theta_4 = 150^\circ$
Eingangs-SIR [dB]	0,02	-2,9	-4,7

**Tabelle 7.5.:** Mischungsparameter.

Mischung	Mix. 13 TU Berlin	Mix. 14 TU Berlin	Mix. 15 TU Berlin
Reverberationszeit $T_R$	190 ms	190 ms	190 ms
Mikrofonabstand $d$	3 cm	3 cm	3 cm
Abtastfrequenz $f_S$	11 kHz	11 kHz	11 kHz
Anzahl der Sprecher $N$	2	3	4
Distanz zwischen dem Sprecher $i$ und dem Arrayzentrum	$L_1 = L_2 = 1,2$ m	$L_1 = L_2 = L_3 = 1,2$ m	$L_1 = L_2 = L_3 = L_4 = 1,2$ m
Angulare Position vom Sprecher $i$ (wie in der Abbildung 7.1)	$\theta_1 = 25^\circ$ $\theta_2 = 70^\circ$	$\theta_1 = 25^\circ$ $\theta_2 = 70^\circ$ $\theta_3 = 120^\circ$	$\theta_1 = 25^\circ$ $\theta_2 = 70^\circ$ $\theta_3 = 120^\circ$ $\theta_4 = 165^\circ$
Eingangs-SIR [dB]	0,05	-3,1	-5,2

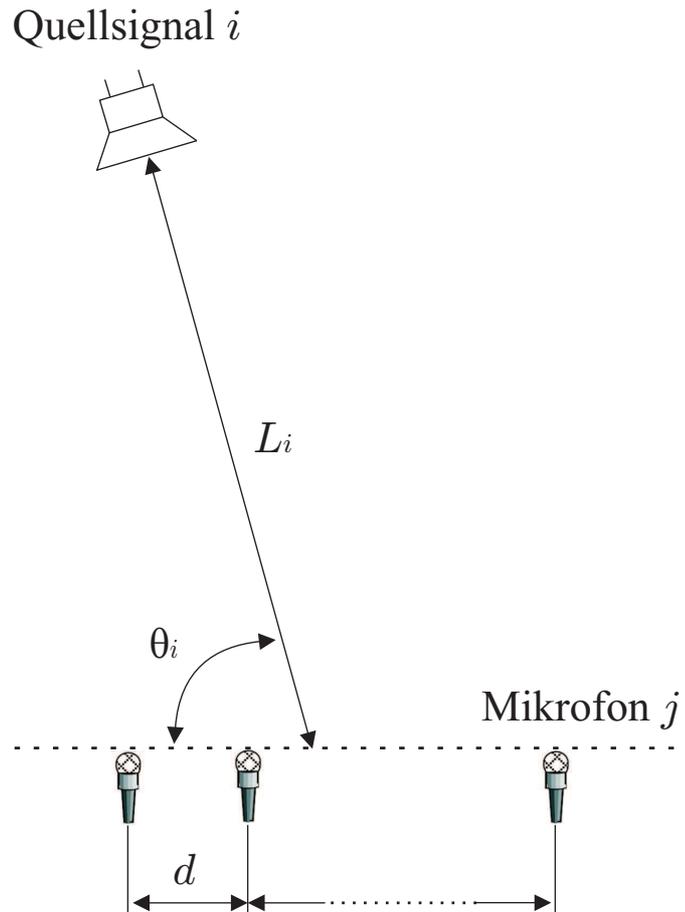
und der Überlappung von 75%. Für die Kurzzeit-Fourier-Transformation wurde das Hanning-Fenster<sup>3</sup> verwendet.

<sup>3</sup>Die Fensterfunktion des Hanning-Fensters ist gegeben durch:

$$w(n) = \frac{1}{2} \left[ 1 + \cos \left( \frac{2\pi n}{M} \right) \right]. \quad (7.1)$$

Dabei ist  $M$  die Fensterbreite und  $n$  ist der Laufindex

$$n \in \left[ -\frac{M}{2}, \dots, \frac{M}{2} - 1 \right]. \quad (7.2)$$



**Abbildung 7.1.:** Schematische Darstellung der Messanordnung. Das Quellensignal  $i$ ,  $i \in 1 \dots N$ , trifft am linearen Mikrofonarray mit dem Einfallswinkel  $\theta_i$  aus der Entfernung  $L_i$  ein und wird vom Mikrofon  $j$ ,  $j \in 1 \dots M$ , aufgenommen.  $d$  ist der Mikrofonabstand im Array.

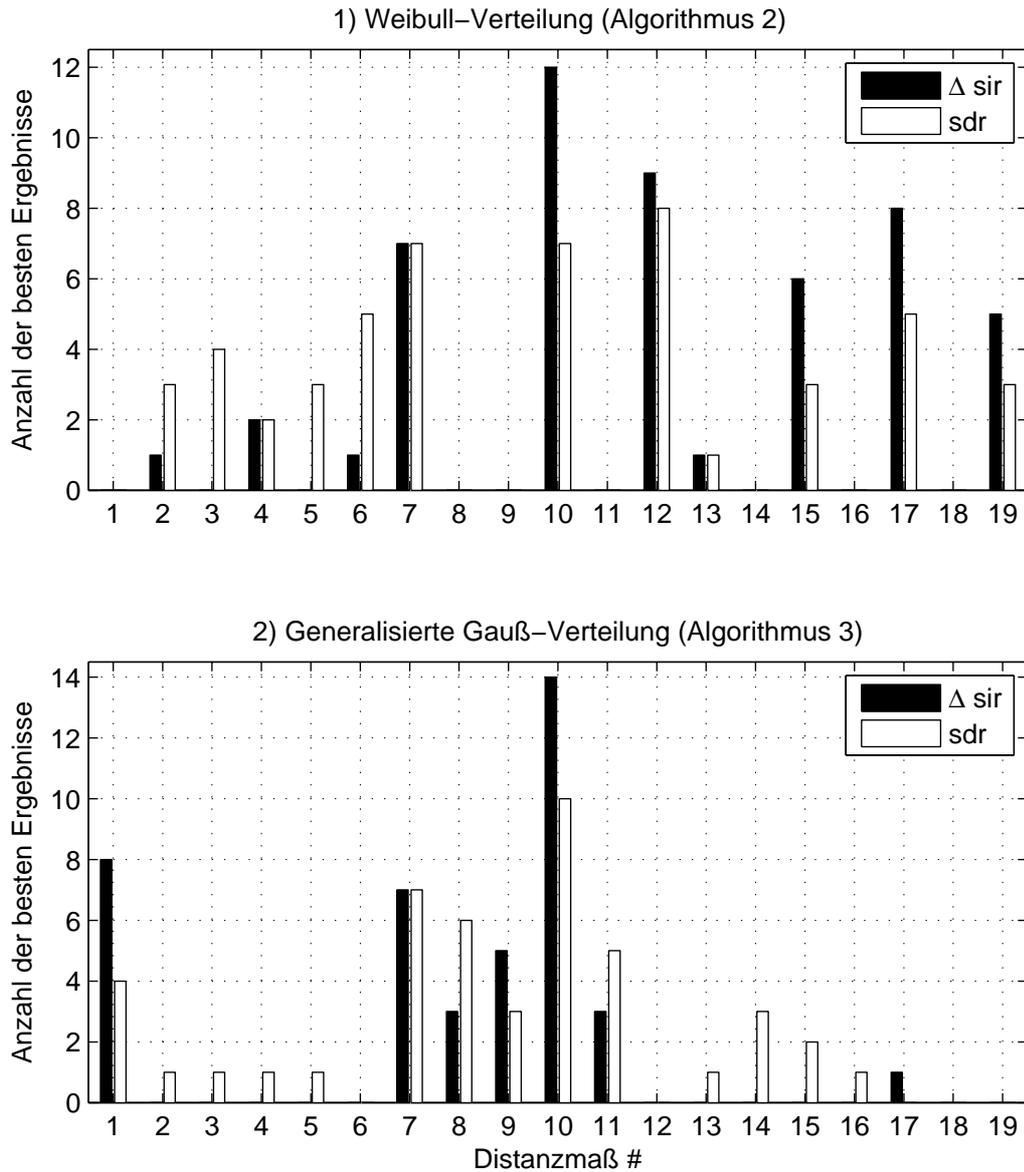
### 7.2.1. Ergebnisse der Permutationskorrekturen mit informationstheoretischen Abstandsmaßen

Die Auswertung der Ergebnisse erfolgte in mehreren Schritten. Als Erstes wurde die Tauglichkeit unterschiedlicher informationstheoretischer Abstandsmaße für die Permutationskorrekturen untersucht. Dazu wurden die Ergebnisse der Permutationskorrekturen basierend auf informationstheoretischen Abstandsmaßen miteinander verglichen. Die beim Vergleich verwendeten Abstandsmaße sind in der Tabelle 7.6 aufgelistet. Die Permutationskorrekturen wurden für die Output-Signale der blinden Quellentrennung im Frequenzbereich für alle Mischungen aus den Tabellen 7.1-7.5 separat ausgehend von den Beträgen der Amplituden unter der Annahme der Weibull-Verteilung (Algorithmus 2) und für die logarithmierten Beträgen der Amplituden unter der Annahme der generalisierten Gauß-Verteilung (Algorithmus 3) berechnet und

ausgewertet. Die Ergebnisse der Auswertung sind in Form eines Histogramms zusammengefasst worden und sind in der Abbildung 7.2 dargestellt.

Wie man sehen kann, wurden die besten Ergebnisse in beiden Fällen mit der Modifizierten Jensen-Tsallis Divergenz mit  $\alpha = 0,5$  (Distanz Nr. 10) erzielt. Außerdem werden die Permutationen im Algorithmus 2 sehr gut mit dem Bhattacharyya-Koeffizienten (Distanz Nr. 12), der Distanz von Chernov (Distanz Nr. 17), der Generalisierten Tsallis-Divergenz mit  $\alpha = 0,5$  (Distanz Nr. 15) und der Modifizierten Jensen-Rényi Divergenz mit  $\alpha = 0,5$  (Distanz Nr. 7) aufgelöst. Bei den logarithmierten Amplituden liegen weitere Maxima der SIR- und SDR-Werte bei der Modifizierten Jensen-Rényi Divergenz mit  $\alpha = 0,5$  (Distanz Nr. 7) und der Gemeinsamen Information (Distanz Nr. 1). Die Tatsache, dass im zweiten Fall die Ergebnisse der Methoden unter der Benutzung der Generalisierten Tsallis- und Generalisierten Rényi-Divergenz und der Distanz von Chernov zurückgeht, ist dadurch zu erklären, dass diese Abstandsmaße sehr empfindlich gegenüber der Modellgenauigkeit sind. So ist die Annahme der Weibull-Verteilung für die Beträge der Amplituden genauer als die Annahme der GGD für die Logarithmen der Amplituden. Andererseits deutet die Tatsache, dass sowohl die Modifizierte Jensen-Tsallis Divergenz als auch die Modifizierte Jensen-Rényi Divergenz in beiden Fällen relativ gute Ergebnisse liefern, auf die Robustheit dieser Distanzen gegenüber der Modell- und Parameterschwankungen hin. Für diese Hypothese sprechen auch die Werte in der Tabelle 7.7. Dort sieht man, dass für die Mittelung über alle Mischungen mit der gleichen Anzahl von Sprechern die Modifizierte Jensen-Tsallis Divergenz mit  $\alpha = 0,5$  (Distanz Nr. 10) im Schnitt die besten Ergebnisse liefert. Die Ergebnisse der Permutationskorrekturen lassen bei allen  $\alpha$ -abhängigen Abstandsmaßen mit dem steigenden  $\alpha$  nach.

Vergleicht man nun die besten Ergebnisse der Mittelungen in der Tabelle 7.7 (fett markierte Zahlen) für die Weibull-Verteilung (linke Seite der Tabelle) mit den Ergebnissen der GGD (rechte Seite der Tabelle), so stellt man fest, dass die Permutationskorrekturen, die auf der generalisierten Gauß-Verteilung basieren, im Schnitt etwas bessere SIR/SDR-Werte liefern. Dies kann dadurch erklärt werden, dass durch die Logarithmierung die Verteilungen der Signale etwas spitzer ausfallen und damit vor allem in den niedrigeren Frequenzbändern besser voneinander zu unterscheiden sind.



**Abbildung 7.2.:** Zusammenstellung der besten Ergebnisse im Sinne von  $\Delta$ SIR (schwarzer Balken) und SDR (weißer Balken) von Algorithmen zur Permutationskorrektur basierend auf informationstheoretischen Abstandsmaßen angewandt auf 1) Beträge der Amplituden der zu korrigierenden Frequenzbänder unter der Annahme der Weibull-Verteilung (siehe Algorithmus 2) und 2) logarithmierten Beträge der Amplituden unter der Annahme der GGD (siehe Algorithmus 3). Die x-Achse bezeichnet die Nummer des Distanzmaßes, so wie sie in der Tabelle 7.6 zu finden ist. Die y-Achse zeigt die Anzahl der besten Ergebnisse über alle Experimente bzw. Anzahl der Fälle, in denen mit dem jeweiligen Distanzmaß das beste Ergebnis erzielt worden ist.

**Tabelle 7.6.:** Verwendete Abstandsmaße.

Nr.	Abstandsmaß
1	Gemeinsame Information
2	Jensen-Rényi Divergenz, $\alpha = 0,5$
3	Jensen-Shannon Divergenz
4	Jensen-Rényi Divergenz, $\alpha = 2$
5	Jensen-Tsallis Divergenz, $\alpha = 0,5$
6	Jensen-Tsallis Divergenz, $\alpha = 2$
7	Modifizierte Jensen-Rényi Divergenz, $\alpha = 0,5$
8	Modifizierte Jensen-Shannon Divergenz
9	Modifizierte Jensen-Rényi Divergenz, $\alpha = 2$
10	Modifizierte Jensen-Tsallis Divergenz, $\alpha = 0,5$
11	Modifizierte Jensen-Tsallis Divergenz, $\alpha = 2$
12	Bhattacharyya-Koeffizient
13	Kullback-Leibler Divergenz
14	Log-Distanz
15	Generalisierte Tsallis-Divergenz, $\alpha = 0,5$
16	Generalisierte Tsallis-Divergenz, $\alpha = 2$
17	Distanz von Chernov, $s = 0,5$
18	Distanz von Kolmogorov, $s = 2$
19	Generalisierte Traubert-Distanz, $n = 8$

### 7.2.2. Ergebnisse der Permutationskorrekturen mit Kernfunktionen

Ähnlich wie im vorigen Abschnitt wurden bei den Tests von Permutationskorrekturen mit der Korrentropie zuerst unterschiedliche Kernfunktionen untersucht. Dazu wurden die Ergebnisse der Permutationkorrekturen basierend auf den Kernfunktionen, so wie in Tabelle 7.8 aufgelistet, berechnet und mit einander verglichen. Die Permutationskorrekturen wurden für die Outputsignale der blinden Quellentrennung im Frequenzbereich für alle Mischungen aus den Tabellen 7.1-7.5 separat ausgehend von den Beträgen der Amplituden und für die logarithmierten Beträge der Amplituden berechnet und ausgewertet. Die Ergebnisse der Auswertung sind in Form von einem Histogramm zusammengefasst worden und sind in der Abbildung 7.3 dargestellt.

Wie man den Histogrammen in der Abbildung 7.3 entnehmen kann, werden die besten Ergebnisse durch den Laplace-Kern (Nr. 3 in der Tabelle 7.8) für die Beträge der Amplituden bzw. durch den polynomialen Kern (Nr. 6 in der Tabelle 7.8) für die logarithmierten Beträge der Amplituden geliefert. Im Gegensatz zu Permutationskorrekturen auf der Basis der informationstheoretischen Abstandsmaßen fallen

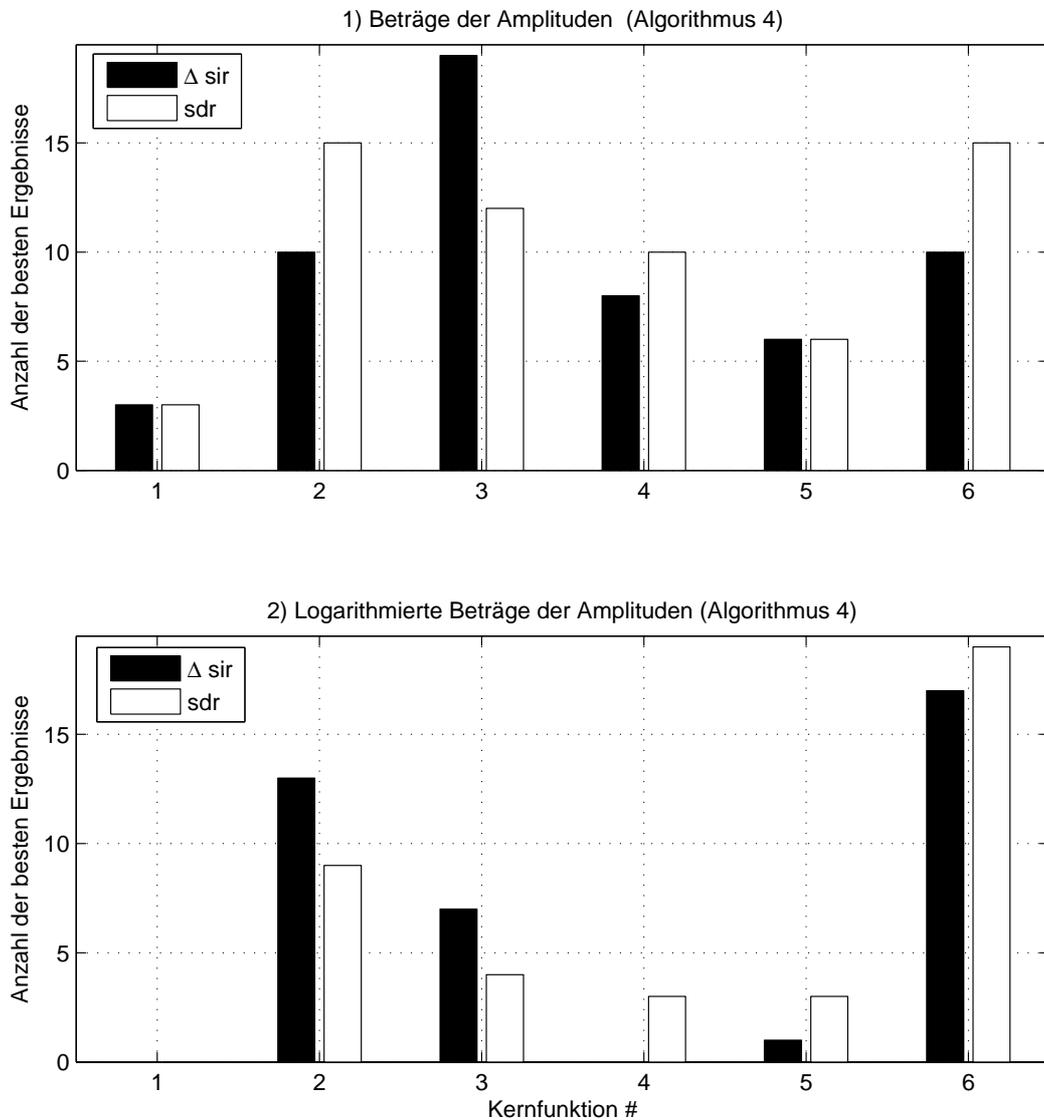
**Tabelle 7.7.:** Mittelung über die berechneten Ergebnisse von  $\Delta$ SIR und SDR-Werten für jedes Distanzmaß (Tabelle 7.6).  $M_i$  steht für den Mittelwert von  $\Delta$ SIR/SDR-Werten berechnet über alle Mischungen von  $i$  Signalen bzw. Sprechern (Tabelle 7.1-7.5).

Abstandsmaß	Algorithmus 2						Algorithmus 3					
	$\Delta$ SIR			SDR			$\Delta$ SIR			SDR		
	M2	M3	M4	M2	M3	M4	M2	M3	M4	M2	M3	M4
1	6.9	5.2	5.8	7.7	4.1	1.6	7.4	7.3	7.7	8.1	5.2	<b>2.4</b>
2	6.2	4.7	6.7	7.4	3.3	1.8	3.6	4.4	5.2	6.0	3.6	1.6
3	0.2	0.8	0.4	7.0	4.0	1.1	3.7	4.6	4.6	6.2	3.8	1.3
4	5.9	5.6	<b>7.3</b>	7.2	3.8	<b>2.1</b>	2.8	3.8	4.9	5.8	3.8	1.4
5	6.3	5.0	6.5	7.5	3.5	1.9	3.8	4.6	5.8	6.3	3.6	1.6
6	6.0	5.5	6.9	7.3	3.8	2.1	2.7	3.4	4.8	5.6	3.7	1.6
7	<b>7.6</b>	<b>6.8</b>	7.1	8.0	<b>5.0</b>	2.0	7.5	7.7	7.9	8.0	5.1	2.3
8	7.4	6.7	7.1	8.0	4.8	2.0	7.5	7.8	<b>8.1</b>	8.0	5.2	2.3
9	1.8	1.1	2.4	4.1	2.5	1.0	7.5	<b>7.9</b>	7.6	8.1	5.2	2.3
10	<b>7.6</b>	5.9	6.0	<b>8.2</b>	4.6	1.8	<b>7.6</b>	<b>7.9</b>	<b>8.1</b>	<b>8.2</b>	<b>5.3</b>	<b>2.4</b>
11	0.7	0.4	2.2	4.4	2.4	0.8	7.3	7.2	7.6	7.9	5.0	2.2
12	7.3	6.5	7.1	8.0	4.8	<b>2.1</b>	2.2	2.7	3.4	5.5	3.3	1.1
13	6.1	5.4	6.5	7.2	4.0	1.7	3.6	4.7	5.0	5.8	3.9	1.5
14	5.4	3.8	5.6	6.8	3.4	1.7	3.4	4.4	3.2	6.1	3.8	1.1
15	6.9	6.5	6.7	7.7	4.7	2.0	3.2	3.5	4.2	5.8	3.6	1.2
16	4.8	3.9	3.2	6.4	3.5	1.1	2.0	2.8	1.5	5.3	3.4	1.0
17	5.4	5.8	6.7	6.5	4.3	<b>2.1</b>	5.4	5.5	6.1	6.8	4.2	1.6
18	2.3	1.6	0.4	5.1	2.8	0.5	2.3	4.0	3.9	4.9	3.2	1.1
19	5.9	6.2	5.6	6.5	4.3	1.8	3.7	5.2	5.6	5.7	3.9	1.5

die Ergebnisse hier für die Beträge der Amplituden und die logarithmierten Beträge der Amplituden etwa gleich aus (Tabelle 7.9). Dies kann dadurch erklärt werden, dass die unterschiedlichen Kernfunktionen etwa die gleichen nichtlinearen Merkmale aus den Amplituden und den Logarithmen der Amplituden extrahieren. Diese werden dann im Merkmalsraum linear verglichen, was für den Ausgleich der Ergebnisse sorgt.

**Tabelle 7.8.:** Verwendete Kernfunktionen.

Nr.	Kernfunktion
1	Gauß-Kern, $\sigma = 1$
2	Epanechnikov-Kern
3	Laplace-Kern
4	Thin Plane Spline Kern, $n = 1$
5	Inverser multiquadratischer Kern, $c = 1$
6	Polynomialer Kern, $d = 2$



**Abbildung 7.3.:** Zusammenstellung der besten Ergebnisse im Sinne von  $\Delta$ SIR (schwarzer Balken) und SDR (weißer Balken) von Algorithmen zur Permutationskorrektur basierend auf Korrentropie angewandt auf 1) Beträge der Amplituden der zu korrigierenden Frequenzbänder und 2) logarithmierten Beträgen der Amplituden (siehe Algorithmus 4). Die x-Achse bezeichnet die Nummer der Kernfunktion, so wie sie in der Tabelle 7.8 zu finden ist. Die y-Achse zeigt die Anzahl der besten Ergebnisse über alle Experimente.

**Tabelle 7.9.:** Mittelung über die berechneten Ergebnisse von  $\Delta$ SIR und SDR-Werten für jede Kernfunktion (Tabelle 7.8).  $M_i$  steht für den Mittelwert von  $\Delta$ SIR/SDR-Werten berechnet über alle Mischungen von  $i$  Signalen bzw. Sprechern (Tabelle 7.1-7.5).

Kern-Fkt.	Algorithmus 4 mit Korrentropie angewandt auf die Amplituden der Signale						Algorithmus 4 mit Korrentropie angewandt auf die logarithmierten Amplituden der Signale					
	$\Delta$ SIR			SDR			$\Delta$ SIR			SDR		
	M2	M3	M4	M2	M3	M4	M2	M3	M4	M2	M3	M4
1	1.0	1.8	1.8	4.4	2.8	0.9	1.0	1.8	1.8	4.4	2.8	0.9
2	7.1	6.1	7.2	7.9	4.5	<b>2.4</b>	7.4	<b>6.7</b>	6.2	8.1	<b>5.1</b>	1.9
3	<b>7.7</b>	<b>6.6</b>	7.0	<b>8.1</b>	<b>5.0</b>	2.0	6.8	6.5	6.1	7.7	4.8	1.6
4	7.5	6.5	<b>7.4</b>	8.0	4.8	2.3	-0.5	0.3	0.3	5.8	4.0	1.3
5	7.5	6.3	<b>7.4</b>	8.0	4.9	2.3	6.2	5.7	4.1	7.4	4.2	1.3
6	7.1	6.1	7.2	7.9	4.5	<b>2.4</b>	<b>7.7</b>	6.5	<b>6.6</b>	<b>8.2</b>	5.0	<b>2.3</b>

### 7.2.3. Vergleich der Algorithmen zur Permutationskorrektur

Im weiteren Schritt wurden die besten vier Algorithmen zur Permutationskorrektur, die in der vorliegenden Arbeit eigenentwickelt wurden, mit den bereits existierenden Algorithmen verglichen. Dazu wurden die 16 bekanntesten Algorithmen aus der Literatur ausgewählt. Diese sind in der Tabelle 7.10 aufgelistet. Dabei ist anzumerken, dass hier bei weitem nicht alle existierenden Algorithmen betrachtet wurden, sondern nur die am weitesten verbreiteten und nur solche, die die Permutationen aufgrund nur eines Kriteriums auflösen, sprich keine kombinierten Kriterien, wie z.B. beim Algorithmus in [MM07] oder [SMAM04]. Wie oben wurden die Algorithmen in Verbindung mit drei ICA-Algorithmen (JADE, FastICA und Inofmax) im Frequenzbereich an allen Mischungen aus den Tabellen 7.1-7.5 getestet und die Ergebnisse in Form von  $\Delta$ SIR und SDR ausgewertet. Die Ergebnisse der Auswertung wurden in Form eines Histogramms zusammengefasst und sind in der Abbildung 7.4 dargestellt.

Wie man leicht sehen kann, zeigt der Vergleich der Algorithmen, dass die besten Ergebnisse bei der Anwendung vom Algorithmus Nr. 2 mit der modifizierten Jensen-Tsallis Divergenz als Abstandsmaß erreicht wurden. Dies wird auch durch die Mittelung über  $\Delta$ SIR bzw. SDR-Werte in der Tabelle 7.11 bestätigt.

Generell ist der Trend, dass die Verfahren, die auf der Konsistenz der Spektren basieren (siehe Abschnitt 4.2.1) z.B. Algorithmen Nr. 2, 3, 8 und 17 in der Tabelle 7.10, im Schnitt etwas besser abschneiden als die Verfahren, die die Konsistenz der Filterkoeffizienten wiederherstellen. Die besten Ergebnisse der Letzteren liefern

die Verfahren Nr. 5 und 18 in der Tabelle 7.10. Diese sind unter anderem durch eine relativ einfache Implementierung und geringere Rechenzeiten (besonders der Algorithmus Nr. 5) gekennzeichnet. Die schlechtere Performance dieser Verfahren gegenüber denen, die auf der Konsistenz der Spektren basieren, kann dadurch erklärt werden, dass diese meistens von dem Mikrofonabstand abhängen und damit bei den größeren Mikrofonabständen (wie z.B. bei den Mischungen aus der Tabelle 7.3) sprich in den Situationen, wo das räumliche Aliasing auftritt, fast komplett versagen. Auch bei den tieferen Frequenzen, wo die Einfallsrichtung des Signals nicht eindeutig bestimmbar ist, besonders bei den Mischungen mit mehr als zwei Sprechern, können die Permutationen von diesen Verfahren nicht richtig aufgelöst werden.

**Tabelle 7.10.:** Verwendete Verfahren zur Permutationskorrektur.

Nr.	Verfahren
1	Algorithmus 2 mit der modifizierten Jensen-Tsallis Divergenz, $\alpha = 0,5$
2	Algorithmus 3 mit der modifizierten Jensen-Tsallis Divergenz, $\alpha = 0,5$
3	Algorithmus 4 mit dem Laplace-Kern angewandt auf die Beträge der Amplituden
4	Algorithmus 4 mit dem Polynomialkern angewandt auf die logarithmierten Beträge der Amplituden
5	Sortierung der Phasendifferenzen der Signale [HKO07a]
6	Clustering der normierten Phasendifferenzen [RAJ07]
7	Kreuzkorrelation von Beträgen der Amplituden [MDC00]
8	Kreuzkorrelation von Leistungsratios [SAM07]
9	Kontinuität der Impulsantwort des Mischungsfilters [PSB03a]
10	Amplitude Modulation Decorrelation [AK00]
11	Kreuzkumulanten von Beträgen der Amplituden [MDC00]
12	Clustering der Amplituden [RAJ07]
13	Glattheit des Mischungsfilters [BKKO01]
14	Minimum der $L_2$ -Norm des Mischungsfilters [SG09]
15	$p$ -Norm der Distanz zwischen den Logarithmen der Beträge von Frequenzbänder mit $p = 2$ [KHK04]
16	Gruppierung der Amplituden nach [PSB03b]
17	Likelihoodratio zwischen den Frequenzbänder [MD03]
18	Basisvektor-Clustering [SAMM06]
19	Beampattern-Minima [BKO03]
20	Cosinus-Distanz zwischen der Elementen der geschätzten Mischungsmatrix [AIO <sup>+</sup> 03]

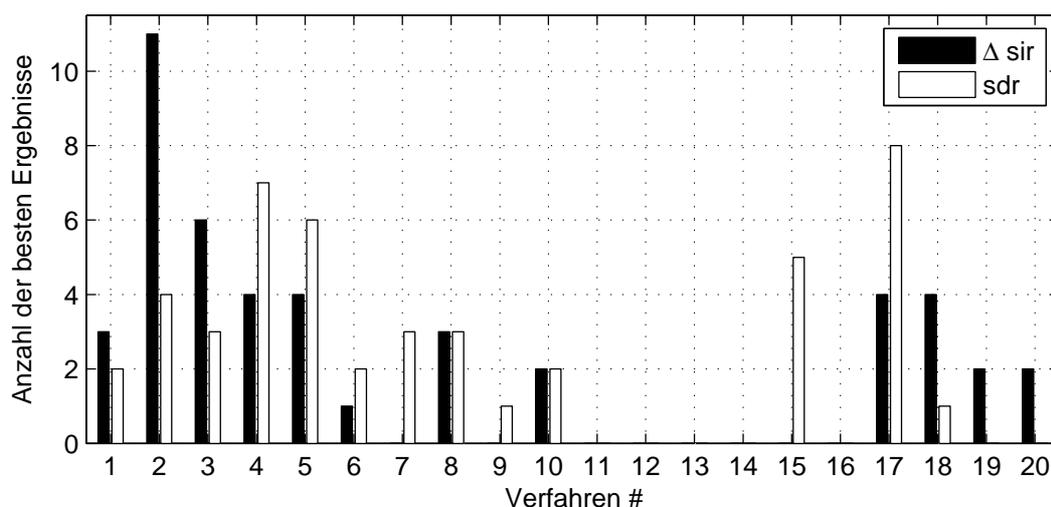


Abbildung 7.4.: Histogramm der besten Ergebnisse im Sinne von  $\Delta$ SIR (schwarzer Balken) und SDR (weißer Balken) von Algorithmen zur Permutationskorrektur. Die x-Achse bezeichnet die Nummer vom Algorithmus, so wie sie in der Tabelle 7.10 zu finden ist. Die y-Achse zeigt die Anzahl der besten Ergebnisse über alle Experimente.

## 7.3. Ergebnisse der Zeit-Frequenz-Maskierung

Im Weiteren werden die Ergebnisse der Zeit-Frequenz-Maskierung vorgestellt. Wegen einer größeren Anzahl der Algorithmen werden die Ergebnisse in mehreren Schritten präsentiert und verglichen. Die in diesem Abschnitt verwendeten Daten (Mischungen aus den Tabellen 7.1-7.5) wurden wie im vorigen Abschnitt in den Frequenzbereich transformiert und mit dem FastICA-Algorithmus verarbeitet<sup>4</sup>. Anschließend wurden auf die getrennten Daten die Algorithmen der Zeit-Frequenz-Maskierung angewandt.

### 7.3.1. Ergebnisse der „klassischen“ Zeit-Frequenz-Maskierung ohne Nachfilterung

Neben dem in dieser Arbeit entwickelten Verfahren (siehe Abschnitt 5.2) wurden weitere Verfahren zur Zeit-Frequenz-Maskierung implementiert und getestet. Es handelt sich dabei um die Algorithmen aus den Abschnitten 5.1.1-5.1.3 nämlich DUET,

<sup>4</sup>Für die Transformation und Trennung wurden dieselben Parameter benutzt, wie im Abschnitt 7.2

**Tabelle 7.11.:** Mittelung über die berechneten Ergebnisse von  $\Delta$ SIR und SDR-Werten für jedes Distanzmaß (Tabelle 7.10).  $M_i$  steht für den Mittelwert der  $\Delta$ SIR/SDR-Werte berechnet über alle Mischungen von  $i$  Signalen bzw. Sprechern (Tabelle 7.1-7.5).

Algorithmus	$\Delta$ SIR			SDR		
	M2	M3	M4	M2	M3	M4
1	7.6	5.9	6.0	8.2	4.6	1.8
2	7.6	<b>7.9</b>	<b>8.1</b>	8.2	<b>5.3</b>	2.4
3	7.7	6.6	7.0	8.1	5.0	2.0
4	7.7	6.5	6.6	8.2	5.0	2.3
5	7.5	6.3	7.8	8.1	4.5	<b>2.6</b>
6	5.1	5.1	5.1	6.6	4.1	1.6
7	7.1	6.1	7.2	7.9	4.5	2.4
8	<b>7.9</b>	6.6	7.8	<b>8.3</b>	4.7	2.5
9	4.0	1.6	2.1	5.8	2.8	1.0
10	6.7	6.4	7.2	7.6	4.5	2.3
11	2.8	2.0	2.5	5.3	2.6	0.7
12	2.4	0.5	1.2	5.3	2.3	0.7
13	0.3	0.7	0.6	4.6	2.4	0.7
14	0.5	1.4	1.3	4.2	3.0	1.0
15	6.3	6.8	7.1	7.4	5.0	2.2
16	7.2	4.8	4.9	8.0	4.2	1.7
17	7.6	6.8	<b>8.1</b>	8.0	<b>5.3</b>	2.4
18	7.3	5.0	6.5	7.7	4.0	1.9
19	7.5	3.3	2.2	7.7	3.0	0.8
20	5.9	4.6	4.7	6.9	4.0	1.8

amplituden-, phasen- sowie die Interferenzen basierte Maskierung. Die implementierten Verfahren mit den für die Auswertung verwendeten Parametern sind in der Tabelle 7.12 zusammengefasst<sup>5</sup>.

In der Tabelle 7.13 sind die Ergebnisse des Vergleichs der einzelnen Verfahren dokumentiert. Die jeweils besten Werte der Maskierung, d.h. die größte SIR-Verbesserung und die kleinste Nutzsignalverzerrung sind fett hervorgehoben.

Die Ergebnisse zeigen, dass bei den gewählten Parametersätzen die besten SIR-Verbesserungen durch die Phasen- und die Interferenzbasierten Maskierungsverfahren erreicht wurden. Dabei funktioniert die interferenzbasierte Maskierung am besten mit den Mischungen mit der kleineren Anzahl von Sprechern. Besonders für die Zweier-Mischungen (Mischungen Nr. 1,4,7,10 und 13) sind hohe SIR-Verbesserungen

<sup>5</sup>Die Vorgehensweise bei dem DUET-Algorithmus wurde an das gemeinsame Schema angepasst, d.h. die mit ICA getrennten Daten wurden anschließend mit einer von DUET berechneten Zeit-Frequenz-Maske multipliziert.

**Tabelle 7.12.:** Verwendete Algorithmen zur Zeit-Frequenz-Maskierung.

Abkürzung	Algorithmus und Parameter
DUET	DUET-Algorithmus
Amp. M.	Amplitudenbasierte Maskierung mit $T = 3$ dB (Gleichung 5.14)
Ph. M.	Phasenbasierte Maskierung mit $g = 20$ und $\theta_T = 0,19\pi$ (Gleichung 5.20)
Int. M. [HKO07a]	Interferenzen basierte Maskierung (Gleichung 5.27) mit $g = 10$ , $R_\Omega = R_\tau = 5$ (Gleichung 5.23) und $P_{\lambda_U} = P_{\lambda_N} = 0,5$ (Gleichung 5.32)

**Tabelle 7.13.:** Ergebnisse der Zeit-Frequenz-Maskierung für die Algorithmen aus der Tabelle 7.12 in Form von mittleren  $\Delta$ SIR/SDR-Werten [dB] über alle Datensätze bzw. Mischungen (Tabellen 7.1-7.5).

Mischung Nr.	ICA	DUET	Amp. M.	Ph. M.	Int. M.
1	8.3/9.4	12.8/ <b>8.3</b>	14.5/8.0	13.4/6.1	<b>22.6</b> /3.7
2	9.1/4.3	13.9/3.7	12.0/ <b>4.3</b>	17.2/2.7	<b>18.8</b> /2.1
3	8.6/1.7	14.7/1.2	10.7/ <b>1.6</b>	<b>17.7</b> /0.5	15.9/0.4
4	4.3/5.6	7.4/ <b>4.5</b>	8.6/4.1	7.6/3.2	<b>12.8</b> /1.8
5	3.4/3.1	6.0/2.1	4.3/ <b>2.9</b>	8.3/1.3	<b>10.3</b> /0.4
6	4.4/0.7	7.7/0.1	4.9/ <b>0.6</b>	<b>10.6</b> /-0.7	7.6/-0.5
7	12.2/11.9	16.9/ <b>10.7</b>	18.0/9.6	19.4/8.5	<b>30.2</b> /4.7
8	12.6/9.3	18.7/7.1	15.4/ <b>8.4</b>	21.5/6.0	<b>27.8</b> /4.1
9	14.3/3.1	21.7/2.4	16.7/ <b>2.8</b>	25.1/1.6	<b>27.5</b> /1.4
10	7.3/6.4	11.5/ <b>5.8</b>	12.8/5.1	11.0/4.4	<b>20.4</b> /2.8
11	9.1/7.0	15.1/5.5	12.2/ <b>6.5</b>	17.9/4.0	<b>19.9</b> /2.8
12	11.5/4.7	17.7/3.4	13.8/ <b>4.5</b>	22.0/2.5	<b>23.2</b> /2.4
13	8.3/8.1	12.3/ <b>7.4</b>	13.6/7.0	12.5/6.2	<b>18.5</b> /5.2
14	5.8/2.7	8.5/ <b>2.5</b>	7.1/ <b>2.5</b>	10.3/2.0	<b>10.8</b> /1.2
15	5.9/1.8	8.2/1.0	7.0/ <b>1.5</b>	<b>11.1</b> /0.5	10.1/0.9

zu vermelden. Hier liegen bei dieser Maskierungsmethode die SIR-Verbesserungen zum Teil um 10 dB über den anderen Maskierungsverfahren. Im Gegensatz dazu ist die phasenbasierte Maskierung besser bei den meisten Vierer-Mischungen. Dies ist dadurch erklärbar, dass bei der steigenden Anzahl von Sprechern (besonders in den Situationen, in denen die meisten Sprecher gleichzeitig aktiv sind) die Sparsität der Quellen abnimmt. Dadurch reduziert sich die Performance der Interferenzbasierten Maskierung, wobei die Trennungsqualität immer noch deutlich besser bleibt als bei DUET oder der Amplitudenbasierten Maskierung.

Das große Problem bei allen Maskierungsverfahren ist, dass die starken SIR-Verbesserungen mit zum Teil sehr hohen Verlusten an Signalqualität verbunden

sind. Daher stellt sich die Frage nach der Wahl eines passenden Parametersatzes, der einen Trade-off zwischen SIR-Verbesserung und Qualitätsverlust gewährleistet. Dies ist aber eine Frage der Anwendung, d.h. derselbe Parametersatz, der gute Ergebnisse bei der Spracherkennung liefert, muss nicht unbedingt die hörbare Qualität verbessern. Es war auch ein Ziel der vorliegenden Arbeit, die Signalzerstörung bei gegebener SIR-Verbesserung zu reduzieren, was zur Einführung von mehrkanaliger Störgeräuschunterdrückung unter Berücksichtigung der Sprachwahrscheinlichkeiten führte, die in folgenden Abschnitten diskutiert wird.

### 7.3.2. Ergebnisse der mehrkanaligen Störgeräuschunterdrückung

Im zweiten Schritt wurden zwei Verfahren der Spektralen Subtraktion (der Maximum-Likelihood-(ML)-Algorithmus, Abschnitt 5.3.2.2, Gleichung 5.75, und der MMSE-Algorithmus, Abschnitt 5.3.2.3, Gleichung 5.86) unter Berücksichtigung der Sprachwahrscheinlichkeiten (siehe Algorithmus 5, Abschnitt 5.4.2) getestet.

Die in diesem Schritt verwendeten Verfahren zur Bestimmung der Sprachwahrscheinlichkeiten sind in der Tabelle 7.14 zusammengefasst.<sup>67</sup> Für die Berechnung von Rauschleistung und der Schätzung der A-Priori und A-Posteriori SNR wurden  $\alpha_D = 0,5$  (Gleichung 5.120) und  $\alpha = 0,95$  (Gleichung 5.121) benutzt. Der Wert von  $G_{\min}$  in der Gleichung 5.101 wurde auf 0,05 gesetzt. Je kleiner  $G_{\min}$  gewählt wird, desto größer ist das SIR aber auch die Verzerrung des Signals.

**Tabelle 7.14.:** Verwendete Methoden zur Bestimmung der Sprachwahrscheinlichkeiten. Für die Berechnung der Sprachwahrscheinlichkeit mit Amplituden-, Phasen- und Interferenzbasierten Masken wurden die selben Parameter benutzt wie in der Tabelle 7.12.

Abkürzung	Algorithmus und Parameter
ohne	keine Sprachwahrscheinlichkeitsberechnung
McAulay	Berechnung der Sprachwahrscheinlichkeit wie in der Gleichung 5.114
Cohen	Berechnung der Sprachwahrscheinlichkeit wie in der Gleichung 5.119
Leist. Ratio	$P(H_1^{(i)}(\Omega, \tau)   \mathbf{Y}(\Omega, \tau)) = \text{powerRatio}_i(\Omega, \tau)$ , Gleichung 4.8
DUET	$P(H_1^{(i)}(\Omega, \tau)   \mathbf{Y}(\Omega, \tau)) = M_i^{\text{DUET}}(\Omega, \tau)$ , Gleichung 5.10
Amp.	$P(H_1^{(i)}(\Omega, \tau)   \mathbf{Y}(\Omega, \tau)) = M_i^{\text{Amp}}(\Omega, \tau)$ , Gleichung 5.13
Phase	$P(H_1^{(i)}(\Omega, \tau)   \mathbf{Y}(\Omega, \tau)) = M_i^{\text{Phase}}(\Omega, \tau)$ , Gleichung 5.20
Interferenz	$P(H_1^{(i)}(\Omega, \tau)   \mathbf{Y}(\Omega, \tau)) = M_i^{\text{Int}}(\Omega, \tau)$ , Gleichung 5.27.

<sup>6</sup>Für die Berechnung der Rauschleistung in dem Fall, dass keine Sprachwahrscheinlichkeitsberechnung durchgeführt wurde, wurde

$$\left(\sigma_D^{(i)}(\Omega, \tau)\right)^2 = \alpha_D \left(\sigma_D^{(i)}(\Omega, \tau - 1)\right)^2 + (1 - \alpha_D) |N_i(\Omega, \tau)|^2 \quad (7.3)$$

benutzt, mit

$$|N_i(\Omega, \tau)| = \max_{j \neq i} (|Y_j(\Omega, \tau)|). \quad (7.4)$$

<sup>7</sup>Für die Berechnung der Sprachwahrscheinlichkeit mit Amplituden-, Phasen- und Interferenzbasierten Masken wurden die selben Parameter benutzt, wie in der Tabelle 7.12.

In den Tabellen 7.15 und 7.16 sind die Ergebnisse für unterschiedliche Arten der Bestimmung der Sprachwahrscheinlichkeiten nach Tabelle 7.14 in Verbindung mit den ML- und der MMSE-Gainfunktionen gegenüber gestellt. Die jeweils besten Resultate der Maskierung, d.h. die größte SIR-Verbesserung und die kleinste Nutzsignalverzerrung sind fett hervorgehoben.

Auch hier ist deutlich zu sehen, dass die besten Ergebnisse der Unterdrückung von Störsprechern in den Fällen erreicht wurden, wo die Phasen- und die Interferenzbasierten Masken zur Bestimmung von Sprachwahrscheinlichkeiten benutzt worden sind. Hier zeigen sich vor allem die Vorteile der Kombination der Verfahren der ein-kanaligen Geräuschreduktion mit den „klassischen“ Maskierungsverfahren, denn die besseren SIR-Werte sind hier im Gegensatz zu den Maskierungsverfahren aus dem Abschnitt 7.3.1 mit den besseren SDR-Werten verbunden. Eine Gegenüberstellung von ML- und MMSE-Algorithmen bei diesem Szenario zeigt, dass die Gain-Funktion von MMSE etwas „aggressiver“ ist und bei dem gleichen Parametersatz bessere SIR-Werte liefert aber auch das Zielsignal stärker verzerrt. Die Trennungsleistung hängt bei den beiden Verfahren unter anderem stark vom Parameter  $G_{min}$  ab. Je kleiner  $G_{min}$  gewählt wird, desto stärker werden die Störsignale unterdrückt. Dies ist in der Abbildung 7.5 gezeigt. Die Abbildung stellt die Abhängigkeit der  $\Delta$ SIR und SDR-Werte von  $G_{min}$  für den ML-Algorithmus mit  $P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau)) = M_i^{\text{Int}}(\Omega, \tau)$  dar. Die Berechnung wurde mit der Mischung Nummer zwei aus der Tabelle 7.1 ( $\Delta$ SIR = 9,1 dB, SDR = 4,3 dB) durchgeführt. Wie man sehen kann, lässt sich durch die geeignete Wahl von  $G_{min}$  ein Trade-off zwischen der Störsprecherunterdrückung und der Nutzsignalverzerrung sehr einfach einstellen.

**Tabelle 7.15.:** Ergebnisse der mehrkanaligen Störgeräuschunterdrückung in Form mittlerer  $\Delta$ SIR/SDR-Werten [dB] für die Arten zur Bestimmung von Sprachwahrscheinlichkeiten aus der Tabelle 7.14 für den Maximum-Likelihood-Algorithmus (Gleichung 5.75) über alle Mischungen (Tabellen 7.1-7.5). Für die Berechnung von Rauschleistung und der Schätzung der a priori und a posteriori SNRs wurden  $\alpha_D = 0,5$  (Gleichung 5.120) und  $\alpha = 0,95$  (Gleichung 5.121) benutzt. Der Wert von  $G_{\min}$  in der Gleichung 5.101 wurde auf 0,05 gesetzt. Je kleiner  $G_{\min}$  gewählt wird, desto größer ist das SIR aber auch die Verzerrung des Signals.

Mischung Nr.	ICA	ohne	McAulay	Cohen	Leist. Ratio
1	8.3/9.4	8.7/ <b>9.2</b>	12.9/7.2	13.4/8.0	14.5/8.3
2	9.1/4.3	9.5/4.0	13.7/2.6	14.1/2.7	14.9/3.8
3	8.6/1.7	8.8/1.3	12.3/0.6	12.6/0.5	14.9/1.3
4	4.3/5.6	4.7/ <b>5.2</b>	7.5/3.4	7.8/3.5	9.3/4.4
5	3.4/3.1	3.8/ <b>3.0</b>	6.2/1.7	6.9/1.7	6.0/2.2
6	4.4/0.7	4.6/ <b>0.6</b>	5.9/0.1	6.2/-0.0	7.4/0.2
7	12.2/11.9	12.7/ <b>11.5</b>	16.1/9.2	16.2/9.9	16.3/9.9
8	12.6/9.3	13.1/ <b>9.0</b>	17.4/6.7	17.6/7.2	18.4/6.9
9	14.3/3.1	14.8/ <b>2.9</b>	19.5/1.8	20.0/1.8	22.0/2.4
10	7.3/6.4	7.8/ <b>6.0</b>	11.6/4.1	12.2/4.1	12.8/5.5
11	9.1/7.0	9.4/ <b>6.7</b>	13.2/4.7	13.6/4.9	16.5/5.6
12	11.5/4.7	12.0/ <b>4.5</b>	16.9/3.1	17.1/3.0	20.2/3.8
13	8.3/8.1	8.7/ <b>7.8</b>	12.7/5.9	13.1/6.5	14.3/7.3
14	5.8/2.7	6.2/ <b>2.6</b>	8.4/2.0	8.7/2.0	8.6/2.4
15	5.9/1.8	5.9/ <b>1.8</b>	6.5/1.5	6.7/1.4	8.3/1.1
Mischung Nr.	ICA	DUET	Amp.	Phase	Interferenz
1	8.3/9.4	14.1/8.2	16.2/7.8	14.8/5.9	<b>23.1</b> /3.9
2	9.1/4.3	15.0/3.6	13.0/ <b>4.2</b>	18.0/2.6	<b>19.8</b> /2.1
3	8.6/1.7	15.8/1.2	11.6/ <b>1.5</b>	<b>18.4</b> /0.5	16.6/0.4
4	4.3/5.6	8.0/4.4	9.8/3.8	8.4/3.2	<b>14.1</b> /2.0
5	3.4/3.1	6.5/2.2	4.6/2.7	9.0/1.3	<b>11.3</b> /0.6
6	4.4/0.7	8.2/0.1	5.2/0.5	<b>11.2</b> /-0.6	8.1/-0.5
7	12.2/11.9	17.5/10.5	18.9/9.2	20.4/8.1	<b>28.6</b> /4.7
8	12.6/9.3	19.4/6.9	16.2/8.1	22.1/5.8	<b>27.7</b> /4.3
9	14.3/3.1	22.5/2.3	17.6/2.7	25.7/1.5	<b>27.9</b> /1.4
10	7.3/6.4	12.6/5.7	14.5/4.9	12.2/4.2	<b>22.0</b> /3.0
11	9.1/7.0	16.3/5.5	13.2/6.2	19.0/3.8	<b>20.8</b> /2.9
12	11.5/4.7	18.9/3.3	14.7/4.3	22.9/2.5	<b>24.0</b> /2.4
13	8.3/8.1	13.7/7.4	15.3/6.6	14.1/6.0	<b>20.1</b> /5.2
14	5.8/2.7	8.9/2.5	7.5/2.4	10.7/2.0	<b>11.6</b> /1.3
15	5.9/1.8	8.6/1.0	7.5/1.4	<b>11.3</b> /0.6	10.7/0.9

**Tabelle 7.16.:** Ergebnisse der mehrkanaligen Störgeräuschunterdrückung in Form mittlerer  $\Delta$ SIR/SDR-Werten [dB] für die Arten zur Bestimmung von Sprachwahrscheinlichkeiten aus der Tabelle 7.14 für den MMSE-Algorithmus (Gleichung 5.86) über alle Mischungen (Tabellen 7.1-7.5). Für die Berechnung von Rauschleistung und der Schätzung der a priori und a posteriori SNRs wurden  $\alpha_D = 0,5$  (Gleichung 5.120) und  $\alpha = 0,95$  (Gleichung 5.121) benutzt. Der Wert von  $G_{\min}$  in der Gleichung 5.101 wurde auf 0,05 gesetzt. Je kleiner  $G_{\min}$  gewählt wird, desto größer ist das SIR aber auch die Verzerrung des Signals.

Mischung Nr.	ICA	ohne	McAulay	Cohen	Leist. Ratio
1	8.3/9.4	9.5/6.5	20.7/4.7	21.3/6.2	21.2/6.6
2	9.1/4.3	11.0/1.5	21.6/1.1	22.0/1.3	19.1/2.7
3	8.6/1.7	10.2/-1.1	18.4/-0.4	19.0/-0.4	19.1/0.9
4	4.3/5.6	5.7/2.5	12.7/1.4	12.9/1.6	13.7/2.7
5	3.4/3.1	6.3/0.3	13.5/0.3	14.8/0.3	7.5/1.4
6	4.4/0.7	7.2/-1.2	11.8/-1.2	12.7/-1.2	10.6/ <b>-0.3</b>
7	12.2/11.9	14.7/8.0	20.3/6.9	20.1/7.8	20.2/7.6
8	12.6/9.3	14.4/4.3	23.7/4.3	23.5/5.1	21.5/5.1
9	14.3/3.1	20.4/0.5	28.4/0.8	28.6/0.9	25.7/1.8
10	7.3/6.4	11.0/2.8	21.0/2.2	21.7/1.9	18.6/4.3
11	9.1/7.0	13.4/2.2	21.6/2.2	21.8/2.6	20.6/4.1
12	11.5/4.7	19.0/1.5	26.0/1.5	25.9/1.4	24.1/2.7
13	8.3/8.1	9.6/5.6	20.9/4.0	21.0/4.8	20.7/5.5
14	5.8/2.7	8.0/0.2	12.6/0.5	13.0/0.5	11.8/1.9
15	5.9/1.8	5.2/-1.1	7.9/ <b>0.6</b>	9.1/0.4	11.2/ <b>0.6</b>
Mischung Nr.	ICA	DUET	Amp.	Phase	Interferenz
1	8.3/9.4	18.7/ <b>7.5</b>	23.1/5.8	21.2/3.9	<b>25.7</b> /2.9
2	9.1/4.3	18.2/2.8	16.9/ <b>3.1</b>	21.5/1.9	<b>23.2</b> /1.7
3	8.6/1.7	18.9/ <b>1.0</b>	15.2/ <b>1.0</b>	<b>21.8</b> /0.1	18.9/0.2
4	4.3/5.6	9.9/ <b>3.8</b>	15.8/1.6	11.9/2.0	<b>18.6</b> /1.5
5	3.4/3.1	8.2/ <b>2.0</b>	5.4/1.2	13.2/0.6	<b>15.7</b> /0.4
6	4.4/0.7	10.0/ <b>-0.3</b>	6.3/ <b>-0.3</b>	<b>14.6</b> /-0.9	10.0/-0.6
7	12.2/11.9	19.5/ <b>9.3</b>	24.2/6.6	25.0/5.3	<b>28.8</b> /3.3
8	12.6/9.3	21.2/5.5	19.6/ <b>6.1</b>	25.1/3.7	<b>29.5</b> /3.1
9	14.3/3.1	25.0/ <b>2.0</b>	22.3/1.8	29.1/0.8	<b>30.3</b> /1.1
10	7.3/6.4	15.9/ <b>5.2</b>	21.8/3.6	17.9/3.0	<b>26.7</b> /2.6
11	9.1/7.0	20.0/ <b>4.7</b>	18.2/4.2	23.7/2.3	<b>24.0</b> /2.1
12	11.5/4.7	22.6/2.6	19.2/ <b>3.4</b>	27.1/1.6	<b>27.5</b> /2.1
13	8.3/8.1	18.5/ <b>6.6</b>	22.9/4.1	20.3/4.8	<b>25.1</b> /4.4
14	5.8/2.7	9.9/ <b>2.2</b>	9.4/1.1	12.6/1.1	<b>14.1</b> /1.0
15	5.9/1.8	10.1/ <b>0.6</b>	10.4/ <b>0.6</b>	12.5/0.4	<b>12.9</b> / <b>0.6</b>

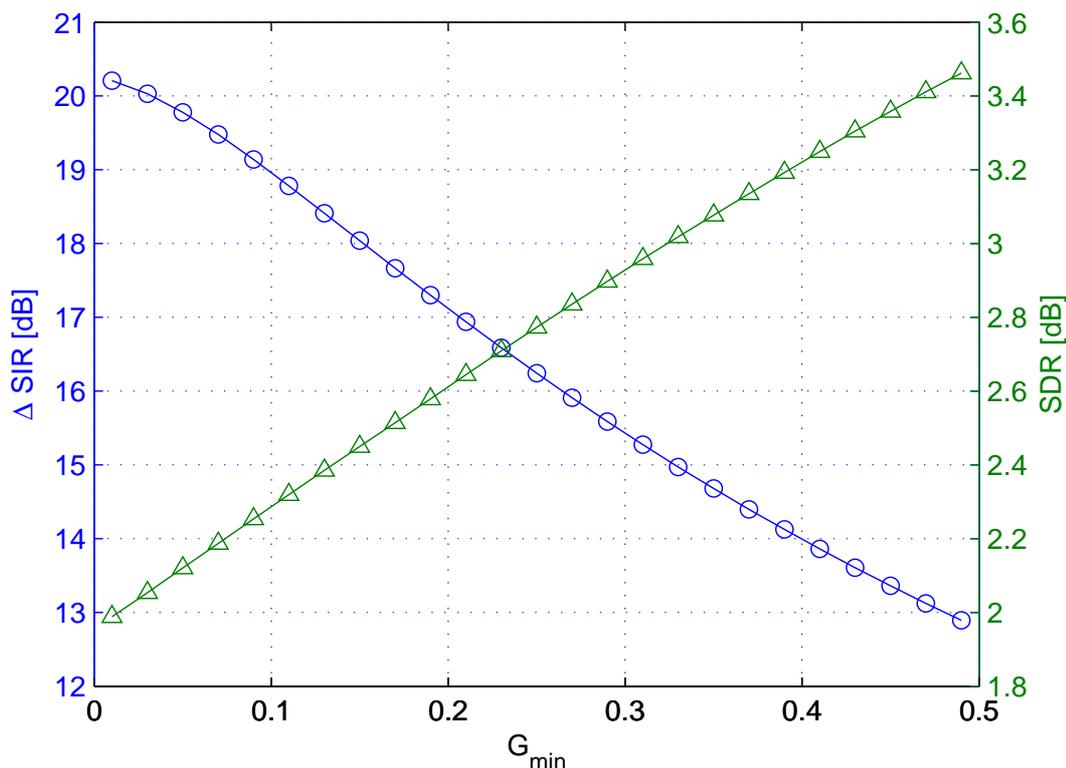


Abbildung 7.5.: Die Abhängigkeit der mittleren  $\Delta SIR$  (o) und SDR-Werte ( $\Delta$ ) von  $G_{min}$  für den ML-Algorithmus mit  $P(H_1^{(i)}(\Omega, \tau) | \mathbf{Y}(\Omega, \tau)) = M_i^{Int}(\Omega, \tau)$  für den Datensatz Nr. 2 aus der Tabelle 7.1 mit FastICA.

### 7.3.3. Subjektive Bewertung der Algorithmen zur Zeit-Frequenz-Maskierung

Neben den objektiven Bewertungen der entwickelten Algorithmen wurden in dieser Arbeit die subjektiven Eindrücke untersucht. Dazu wurden in mehreren Hörtests die unterschiedlichen Verfahren zur Zeit-Frequenz-Maskierung untersucht und miteinander verglichen. Der Vergleich erfolgte mit dem subjektiven Bewertungsverfahren MUSHRAM [Vin05] gemäß Abschnitt 6.3. Für die Tests wurden zwei Sprachsignale (Signal Nr. 3 aus der Mischung Nr. 2 und Signal Nr. 1 aus der Mischung Nr. 8) verwendet, die wie in den oberen Abschnitten mit FastICA und unterschiedlichen Maskierungsverfahren verarbeitet wurden. Als Anker wurden zum Ersten das Referenzsignal tiefpassgefiltert mit der Grenzfrequenz 3500 Hz und ein Kanal der ungetrennten Mischung verwendet [Vin05]. Die für die Hörtests verwendeten Maskierungsverfahren sind in der Tabelle 7.17 zusammengefasst.

Die verarbeiteten Signale wurden von 22 Personen im Alter von 22 bis 43 Jahren bewertet. Die Ergebnisse der subjektiven Bewertungen sind in den Abbildungen 7.6-7.8

**Tabelle 7.17.:** Die in den Hörtests verglichenen Algorithmen zur Nachverarbeitung der Sprachsignale nach der blinden Quellentrennung.

Abkürzung	Algorithmus und Parameter
ICA	Keine Maskierung angewandt
ML-Ph	ML-Algorithmus mit der Phasenbasierten Maskierung
ML-Int	ML-Algorithmus mit der Interferenzen basierten Maskierung
Ph. M.	Phasenbasierte Maskierung
Int. M.	Interferenzen basierte Maskierung
Amp. M.	Amplitudenbasierte Maskierung
DUET	DUET-Algorithmus

in Form von *Box-Whisker-Plots*<sup>8</sup> zu sehen. In der Tabelle 7.18 sind die Mittelwerte der subjektiven Bewertungen für die drei getesteten Signaleigenschaften über alle getesteten Algorithmen zusammengefasst. Die besten Werte der Mittelung sowie die größten  $\Delta$ SIR/SDR-Werte sind fett hervorgehoben.

Bedenkt man, dass die Sprachverständlichkeit wahrscheinlich das wichtigste Kriterium ist [Pet02], so zeigt sich, dass man mit der Nachverarbeitung mit den Methoden der mehrkanaligen Störgeräuschunterdrückung mit der Schätzung von Sprachwahrscheinlichkeiten stets gute Ergebnisse erreicht. Vor allem das Verfahren ML-Int (Tabelle 7.17) erzielt die besten Bewertungen. Dies gilt auch für die Unterdrückung der Störsprecher. Dies ist vor allem dadurch zu erklären, dass trotz Signalverzerrungen, die mit diesen Verfahren verarbeiteten Signale natürlicher klingen, als nach der gewöhnlichen Zeit-Frequenz-Maskierung. Die Einschätzung der Unterdrückung der Störsprecher von ICA ohne Maskierung korrespondiert im übrigen mit den Ergebnissen der subjektiven Bewertungen, die von M.G. Jafari et al. mit MUSHRAM durchgeführt wurden und die in [JVA<sup>+</sup>08] zu finden sind. Wie in der vorliegenden Arbeit wird die Performance von ICA dort von den Testpersonen auf 30 bis 35 eingeschätzt. Hier liegt sie bei 34,9 (Tabelle 7.18).

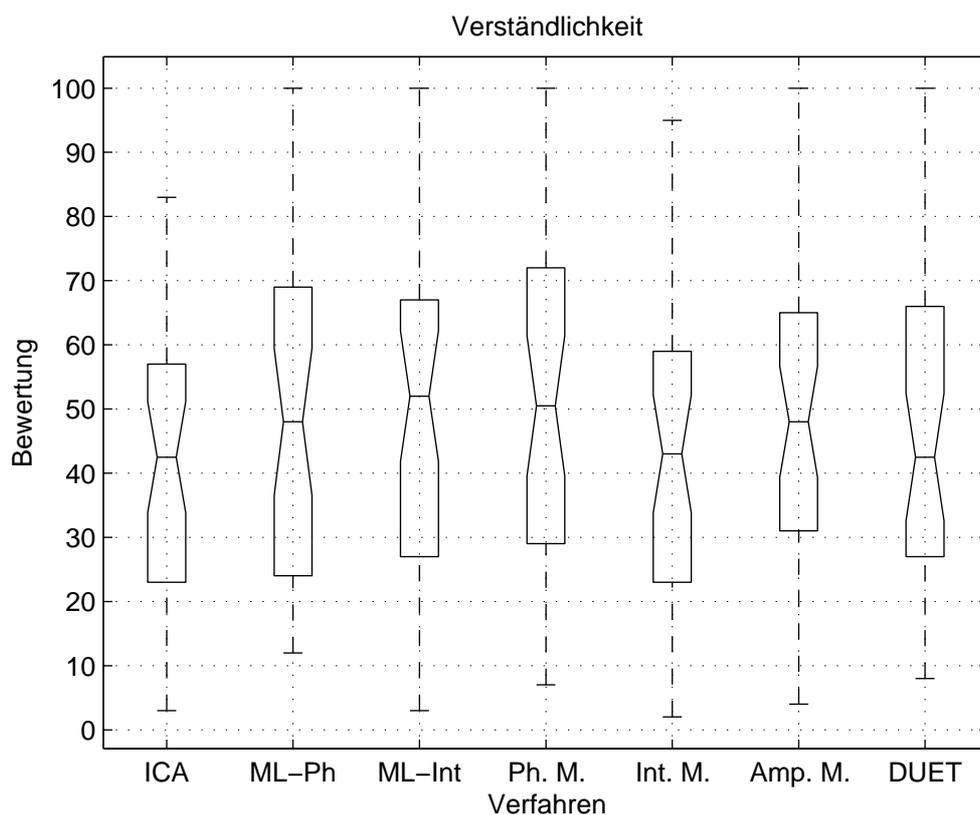
Besonders auffällig ist, dass trotz schlechterer SDR-Werte die Signalverzerrung von Amp. M. von der Hörerschaft höher eingeschätzt wird als bei ICD ohne Maskierung. Eine Ursache dafür könnte darin liegen, dass bei der Einschätzung der Verzerrung auch die bleibenden Interferenzen eine Rolle spielen. Andererseits ist die Verzerrung

---

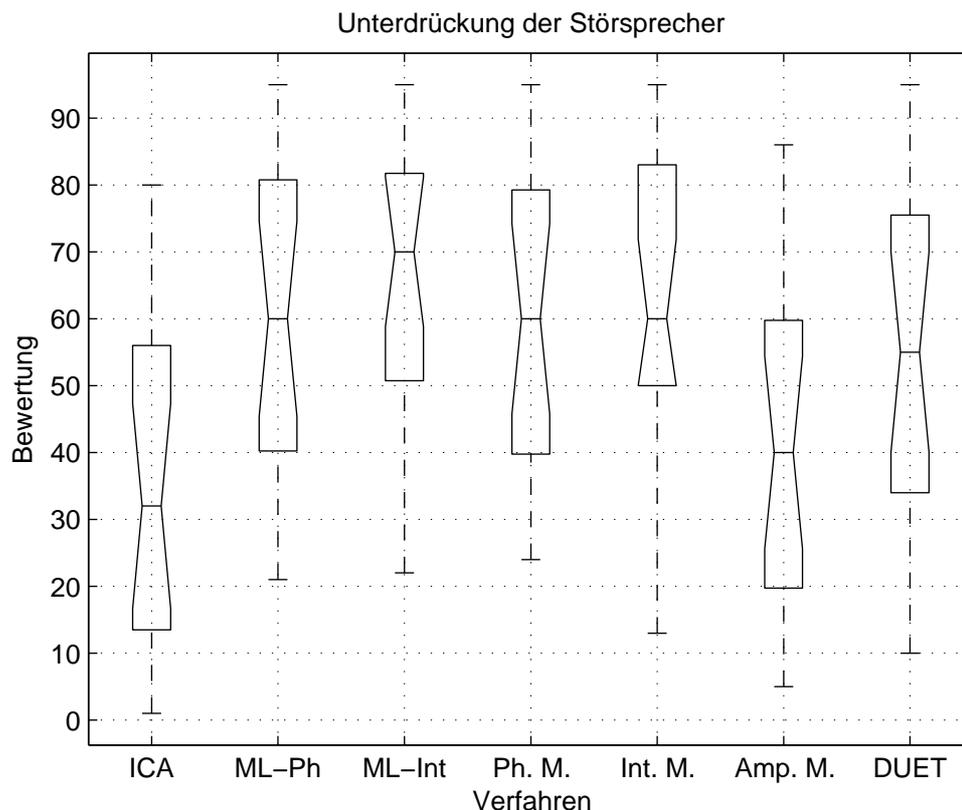
<sup>8</sup>Box-Whisker-Plot ist eine grafische Darstellung einer statistischen Verteilung. Es wird durch sechs Streuungs- und Lagemaße (der Median, die zwei Quartile, die beiden Extremwerte und die Ausreißer) festgelegt. Die gestrichelte Linie im Diagramm zeigt dabei den Bereich in der Verteilung, in dem keine Ausreißer vorkommen. Die Box entspricht dem Bereich, in dem die mittleren 50% (zwischen 25% und 75%) der Daten liegen. Sie wird also durch das obere und das untere Quartil begrenzt. Und der Median teilt das gesamte Diagramm in zwei Hälften, in denen jeweils 50% der Daten liegen [Tuk].

bei den beiden ML-Verfahren kleiner als bei den entsprechenden gewöhnlichen Maskierungsverfahren (*Ph. M.* und *Int. M.*). Bedenkt man, dass bei diesen Verfahren die bessere Sprachverständlichkeit und bei dem *ML-Int* die bessere Störsprecherunterdrückung sowie bessere SIRs und SIR/SDR-Verhältnisse bei den beiden erreicht wurden, so kann man sagen, dass die Verfahren der mehrkanaligen Störgeräuschunterdrückung eine bessere Wahl für die Nachfilterung der getrennten Sprachsignale bilden.

Zusammenfassend lässt sich festhalten, dass sich die Verwendung der mehrkanaligen Störgeräuschunterdrückung mit der Schätzung von Sprachwahrscheinlichkeiten als Nachverarbeitung für die blinde Quellentrennung im Frequenzbereich als vorteilhaft erwiesen hat. Dies wird sowohl durch die Verbesserung des Signal-zu-Interferenz-Verhältnisses sowie durch die durchgeführten Hörtests bestätigt.



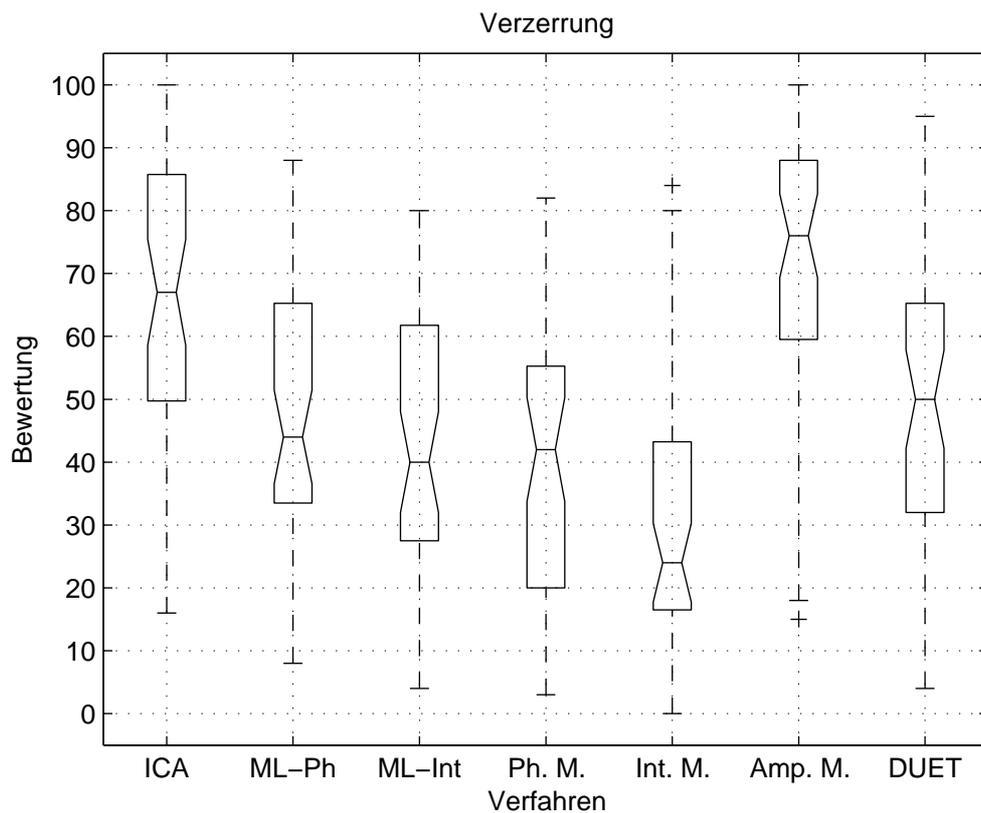
**Abbildung 7.6.:** Ergebnisse der subjektiven Bewertungen der Sprachverständlichkeit nach der Zeit-Frequenz-Maskierung (Tabelle 7.17) gemittelt über 22 Testpersonen für zwei Testsignale. Größere Werte in den Bewertungen entsprechen einer besseren Sprachverständlichkeit.



**Abbildung 7.7.:** Ergebnisse der subjektiven Bewertungen der Unterdrückung der Störsprecher durch die Zeit-Frequenz-Maskierung (Tabelle 7.17) gemittelt über 22 Testpersonen für zwei Testsignale. Größere Werte in den Bewertungen entsprechen einer besseren Unterdrückung der Störsprecher durch die Algorithmen.

**Tabelle 7.18.:**  $\Delta$ SIR/SDR-Werte für die bei den Hörtests verwendeten Signale sowie die Mittelwerte der subjektiven Bewertungen für die drei getesteten Signaleigenschaften über alle getesteten Algorithmen.

Verfahren	$\Delta$ SIR/SDR Signal 1 [dB]	$\Delta$ SIR/SDR Signal 2 [dB]	Verständ- lichkeit	Störsprecher- unterdrückung	Verzerrungs- freiheit
ICA	10,5/ <b>4,6</b>	15,4/ <b>8,5</b>	41,8	34,9	66,5
ML-Ph	18,6/3,1	27,5/5,4	50,5	59	48,3
ML-Int	<b>25,3</b> /2,2	33,5/3,8	<b>51,3</b>	<b>64,3</b>	43,8
Ph. M.	17,4/3,2	27,1/5,5	49,1	60,8	40,6
Int. M.	24,3/2,2	<b>34,5</b> /3,5	41,9	62,3	31,0
Amp. M.	13,1/4,2	19,4/7,8	49,9	41,2	<b>71,3</b>
DUET	15,3/3,3	23,5/6,7	47,8	52,7	49,0



**Abbildung 7.8.:** Ergebnisse der subjektiven Bewertungen der Verzerrungsfreiheit nach der Zeit-Frequenz-Maskierung (Tabelle 7.17) gemittelt über 22 Testpersonen für zwei Testsignale. Kleinere Werte in den Bewertungen entsprechen einer größeren Verzerrung, die im Signal durch die Algorithmen verursacht wurde.



## 8. Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurden zwei Themengebiete der blinden Quellentrennung konvolutiver Mischungen von Sprachsignalen bearbeitet. Zum Einen wurde die Problematik der Permutationskorrekturen betrachtet, die für die blinde Quellentrennung im Frequenzbereich beim heutigen Stand der Forschung unumgänglich sind. Zum anderen wurden die Methoden der Zeit-Frequenz-Maskierung diskutiert, die optional für die Verbesserung der Trennungsleistung der blinden Quellentrennung als Nachverarbeitungsstufe eingesetzt werden können.

Nach der Vorstellung verschiedener Verfahren zur Permutationskorrektur wurde auf Vergleichsmöglichkeiten der zu korrigierenden Frequenzbänder auf der Basis informationstheoretischer Abstandsmaße sowie der Kernfunktionen eingegangen. Ausgehend davon wurden zwei Klassen von Algorithmen zur Permutationskorrektur entwickelt: die Informationstheoretischen und die Kernfunktionenbasierten. Die Ersten gehen von bestimmten Annahmen über die Verteilung der Signale im Frequenzbereich aus. Zur Lösung des Permutationsproblems auf diesem Weg wurde ein neues Abstandsmaß definiert, das die Eigenschaften von Rényi- und Tsallis-Entropien ausnutzt und dessen Gültigkeit im Kapitel 4 gezeigt wurde. Die zweite Klasse von Algorithmen geht von der Annahme der linearen Trennbarkeit der Signaleigenschaften im höherdimensionalen Merkmalsraum aus, welcher mit Hilfe der Kernfunktionen erreicht wird, und benutzt die Korrentropie zum Finden der richtigen Reihenfolge der Signale. Beide Klassen von Algorithmen sind sowohl auf die Beträge der Amplituden oder auf die logarithmierten Beträge der Amplituden der zu korrigierenden Frequenzbänder anwendbar. Anschließend wurden die neu entwickelten Verfahren mit den realen Raumaufnahmen getestet und mit den bereits existierenden Verfahren verglichen. Der Vergleich zeigt die Robustheit der in dieser Arbeit entwickelten Algorithmen. Wegen der getroffenen Annahmen bezüglich der statistischen Verteilung der Sprachsignale haben die informationstheoretischen Verfahren besonders gute Ergebnisse gezeigt. Dies ist im Kapitel 7 dokumentiert.

Der zweite Teil dieser Arbeit widmet sich dem Problem der Signalverbesserung nach

der blinden Quellentrennung im Frequenzbereich. Dazu wurde zuerst ein Algorithmus entwickelt, der basierend auf der Annahme der zeitlichen Dominanz eines Signals pro Zeit-Frequenz-Schritt eine Maskenfunktion zur Reduktion der verbliebenen Interferenzen definiert. Die Maskenfunktion kann sowohl binär als auch weich definiert werden. Der entwickelte Algorithmus wurde zusammen mit den bekannten Zeit-Frequenz-Maskierungsverfahren an mehreren realen Raumaufnahmen getestet. Die Ergebnisse zeigen deutliche SIR-Verbesserungen gegenüber bloßer ICA. In den Fällen, in denen weniger als vier Sprecher in der Mischung aktiv waren, hat der neue Algorithmus die besten Ergebnisse geliefert. Mit steigender Anzahl von gleichzeitig sprechenden Personen bzw. der gleichzeitig aktiven Quellen verringert sich die Sparsität der Signale und somit lässt die Performance des Verfahrens nach. An dieser Stelle ist jedoch anzumerken, dass Situationen, in denen über vier Zielsprecher gleichzeitig aktiv sind und sich in dem ungefähr gleichen Abstand zum Mikrofonarray befinden, eher unrealistisch sind.

Im Weiteren wurde eine Kombination der einkanaligen Verfahren der Spektralen Subtraktion zur Geräuschreduktion mit der mehrkanaligen Zeit-Frequenz-Maskierung vorgeschlagen. Dabei wurde das Ziel verfolgt, die Signalverzerrung und die mit der Zeit-Frequenz-Maskierung verbundenen Artefakte (z.B. Musikal Tones) zu reduzieren ohne die Trennungsleistung zu verschlechtern. Die Hauptidee war dabei, die Wahrscheinlichkeiten für die Sprachaktivität bei der Spektralen Subtraktion aus den mehrkanaligen Informationen zu gewinnen. Die Wirksamkeit dieses Ansatzes wurde anhand praktischer Versuche verifiziert. Die dabei entstandenen Verfahren zeigen eine hohe SIR-Verbesserung (in den meisten Fällen eine Höhere als bei den „klassischen“ Zeit-Frequenz-Maskierung). Im Vergleich zur Zeit-Frequenz-Maskierung ist diese SIR-Verbesserung mit besseren SDR-Werten verbunden. Eine weitere Bestätigung für die bessere Qualität der resultierenden Signale ergab sich durch Psychoakustische Tests. Die Ergebnisse der Hörtests zeigen, dass die Verfahren der mehrkanaligen Störgeräuschunterdrückung unter Berücksichtigung der Sprachwahrscheinlichkeiten bessere Sprachverständlichkeit und Störsprecherunterdrückung liefern. Dabei bleibt die von den Testpersonen wahrgenommene Signalverzerrung kleiner als bei der zugrunde liegenden „klassischen“ Zeit-Frequenz-Maskierung. Dies erlaubt den Einsatz der hier vorgestellten Verfahren zur Signalverbesserung für die akustischen Mensch-Maschine-Schnittstellen, Mobiltelefone oder Hörgeräte.

Als Ausblick sei angemerkt, dass die in der vorliegenden Arbeit präsentierten Algorithmen der Spektralen Subtraktion mit der mehrkanaligen Schätzung der Wahrscheinlichkeiten für die Sprachaktivität auch auf dem Gebiet der robusten Spracherkennung vielversprechende Ergebnisse zeigen. Wie in [HKO07b] gezeigt wurde, steigen die Erkennungsraten beim Einsatz dieser Verfahren in der Spracherkennung.

---

Dieser Zweig wurde aber in der vorliegenden Arbeit nicht weiter verfolgt, da dies einer umfangreichen Erläuterung der Thematik der Spracherkennung bedürfte, was den Rahmen dieser Arbeit sprengen würde.

Des Weiteren wäre es interessant, die durch die Zeit-Frequenz-Maskierung gewonnenen Informationen in Verbindung mit psychoakustischen Masken zu testen bzw. diese bei der Berechnung einer Psychoakustischen Maske im mehrkanaligen Fall zu integrieren. Bis jetzt wurden die psychoakustischen Masken nur in einkanaligen Szenarien betrachtet [GJV98, Loi07, Pet02, KK02b]. Der Vorteil der psychoakustischen Masken besteht darin, dass die Maskenfunktion die Eigenschaften des menschlichen Gehörs berücksichtigt und somit eine bessere hörbare Qualität sowie eine bessere Sprachverständlichkeit erwarten lässt.



# A. Ausdrücke für die Shannon-, Rényi- und Tsallis-Entropien bei Vorliegen der Weibull- und der generalisierten Gauß-Verteilungen

Für die Berechnung der informationstheoretischen Abstandsmaße werden die Ausdrücke für die Entropien benötigt. Eine Möglichkeit ist es, die Entropien numerisch aus der Mittelung der Verteilungsdichte zu bestimmen. Diese gibt allerdings meistens eine schlechtere Schätzung der Entropie als eine geschlossene Form. Im Weiteren werden die Ausdrücke für die Shannon-, Rényi- und Tsallis-Entropien bei Vorliegen der Weibull- und der generalisierten Gauß-Verteilungen angegeben.

## A.1. Entropien der Weibull-Verteilung

Die Rényi-Entropie der Weibull-Verteilung wurde für die vorliegende Arbeit mit Hilfe des Online-Integrationstools von Mathematica<sup>1</sup>, das für die Lösung von symbolischen Integralen programmiert ist, dargestellt.

Die Rényi-Entropie der Weibull-Verteilung ist gegeben durch

$$H_{\alpha}^R(X) = \frac{1}{1-\alpha} \log \left( \int_{\mathbb{X}} f_X^{\alpha}(x) dx \right) \quad (\text{A.1})$$

$$= \frac{1}{1-\alpha} \log \left( \left( \frac{\lambda}{k} \right)^{1-\alpha} A \Gamma \left( \alpha - \frac{\alpha}{k} + \frac{1}{k} \right) \right) \quad (\text{A.2})$$

$$= \log \left( \frac{\lambda}{k} \right) + \frac{1}{1-\alpha} \log A + \frac{1}{1-\alpha} \log \left( \Gamma \left( \alpha - \frac{\alpha}{k} + \frac{1}{k} \right) \right). \quad (\text{A.3})$$

---

<sup>1</sup><http://integrals.wolfram.com/index.jsp>

mit

$$A = \alpha^{(-\alpha + \frac{\alpha}{k} - \frac{1}{k})} \quad (\text{A.4})$$

und  $\Gamma(\cdot)$  als Gammafunktion (Gleichung 2.22). Wählt man  $\alpha \rightarrow 1$ , so ergibt sich bei der Anwendung von der Regel von L'Hospital

$$\lim_{\alpha \rightarrow 1} \frac{1}{1 - \alpha} \log A = \lim_{\alpha \rightarrow 1} \frac{(-\alpha + \frac{\alpha}{k} - \frac{1}{k}) \log \alpha}{1 - \alpha} \quad (\text{A.5})$$

$$= \lim_{\alpha \rightarrow 1} \frac{\frac{d}{d\alpha} ((-\alpha + \frac{\alpha}{k} - \frac{1}{k}) \log \alpha)}{-1} \quad (\text{A.6})$$

$$= \lim_{\alpha \rightarrow 1} - \left( \left( -1 + \frac{1}{k} \right) \underbrace{\log \alpha}_{\rightarrow 0} + \frac{(-\alpha + \frac{\alpha}{k} - \frac{1}{k})}{\alpha} \right) \quad (\text{A.7})$$

$$= - \frac{(-1 + \frac{1}{k} - \frac{1}{k})}{1} = 1 \quad (\text{A.8})$$

und

$$\lim_{\alpha \rightarrow 1} \frac{1}{1 - \alpha} \log \left( \Gamma \left( \alpha - \frac{\alpha}{k} + \frac{1}{k} \right) \right) = \lim_{\alpha \rightarrow 1} \frac{\frac{d}{d\alpha} \log \left( \Gamma \left( \alpha - \frac{\alpha}{k} + \frac{1}{k} \right) \right)}{\frac{d}{d\alpha} (1 - \alpha)}. \quad (\text{A.9})$$

Mit

$$\Psi(z) = \frac{d}{dz} \log(\Gamma(z)) \quad (\text{A.10})$$

als die Digamma-Funktion Damit wird die Gleichung A.9 zu

$$\begin{aligned} \lim_{\alpha \rightarrow 1} \frac{1}{1 - \alpha} \log \left( \Gamma \left( \alpha - \frac{\alpha}{k} + \frac{1}{k} \right) \right) &= \lim_{\alpha \rightarrow 1} \frac{\Psi \left( \alpha - \frac{\alpha}{k} + \frac{1}{k} \right)}{-1} \\ &\quad \times \frac{d}{d\alpha} \left( \alpha - \frac{\alpha}{k} + \frac{1}{k} \right) \end{aligned} \quad (\text{A.11})$$

$$= \lim_{\alpha \rightarrow 1} \frac{\Psi \left( \alpha - \frac{\alpha}{k} + \frac{1}{k} \right)}{-1} \left( 1 - \frac{1}{k} \right) \quad (\text{A.12})$$

$$= -\Psi(1) \left( 1 - \frac{1}{k} \right). \quad (\text{A.13})$$

Damit erhält man einen Ausdruck für die Shannon-Entropie

$$H_1(X) = \log \left( \frac{\lambda}{k} \right) + 1 + \gamma \left( 1 - \frac{1}{k} \right) \quad (\text{A.14})$$

mit  $\gamma = -\Psi(1) \approx 0,57721566$  als Euler-Mascheroni-Konstante.

Die Tsallis-Entropie ergibt sich aus der Gleichung A.2 und 2.62

$$H_q^T(X) = \frac{1}{q-1} \left( 1 - \left( \frac{\lambda}{k} \right)^{1-q} A\Gamma \left( q - \frac{q}{k} + \frac{1}{k} \right) \right). \quad (\text{A.15})$$

Die Shannon-Entropie der multivariaten Weibull-Verteilung (Gleichung 2.33) ist gegeben durch [DV00, ZN05]

$$H_1(\mathbf{X}) = - \sum_{i=1}^N \log \left( \frac{\beta + i - 1}{\lambda_i} \right) + (\beta + N) \sum_{i=1}^N \frac{1}{\beta + i - 1} - \frac{N}{\beta} - \sum_{i=1}^N \log k_i + (\gamma + \log \beta) \sum_{i=1}^N \frac{k_i - 1}{k_i}. \quad (\text{A.16})$$

## A.2. Entropien der Generalisierten Gauß-Verteilung

Die Rényi-Entropie der Generalisierten Gauß-Verteilung ist gegeben durch

$$H_\alpha^R(X) = \frac{1}{1-\alpha} \log \left( \int_{\mathbb{X}} f_X^\alpha(x) dx \right) \quad (\text{A.17})$$

$$= \frac{1}{1-\alpha} \log \alpha^{-1/\beta_X} - \log \left( \frac{\beta_X}{2a\Gamma(1/\beta_X)} \right). \quad (\text{A.18})$$

Da für  $\alpha \rightarrow 1$  gilt

$$\lim_{\alpha \rightarrow 1} \frac{1}{1-\alpha} \log \alpha = \lim_{\alpha \rightarrow 1} -\frac{1}{\alpha} = -1, \quad (\text{A.19})$$

wird die Shannon-Entropie der GGD zu

$$H_1(X) = \frac{1}{\beta_X} - \log \left( \frac{\beta_X}{2a\Gamma(1/\beta_X)} \right). \quad (\text{A.20})$$

Den Ausdruck für die Tsallis-Entropie erhält man nach dem Einsätzen der Gleichung A.18 in die Gleichung 2.62 und Umformen nach  $H_q^T(X)$

$$H_q^T(X) = \frac{1}{q-1} \left( 1 - \left( \frac{\beta_X}{2a\Gamma(1/\beta_X)} \right)^{q-1} q^{-1/\beta_X} \right). \quad (\text{A.21})$$

Die Shannon-Entropie der multivariaten GGD (Gleichung 2.24) ist gegeben

durch

$$H_1(\mathbf{X}) = \frac{1}{2} \log \left( \det\{\boldsymbol{\Sigma}\} \prod_{i=1}^N \left[ \frac{\beta_{\tilde{x}_i}}{(2a_{\tilde{x}_i} \Gamma(1/\beta_{\tilde{x}_i}))} \exp \left\{ -\frac{1}{\beta_{\tilde{x}_i}} \right\} \right] \right). \quad (\text{A.22})$$

# Literaturverzeichnis

- [ABR64] AIZERMAN, A. ; BRAVERMAN, E. M. ; ROZONOER, L. I.:  
Theoretical foundations of the potential function method in pattern  
recognition learning. In: *Automat. Remote Contr.* 25 (1964), S.  
917–936
- [AIO<sup>+</sup>03] ASANO, F. ; IKEDA, S. ; OGAWA, M. ; ASOH, H. ; KITAWAKI, N.:  
Combined approach of array processing and independent  
component analysis for blind separation of acoustic signal. In:  
*IEEE Transactions on Speech and Audio Processing* 11 (2003), Mai,  
Nr. 3, S. 2104–2215
- [AK00] ANEMÜLLER, J. ; KOLLMEIER, B.: Amplitude modulation  
decorrelation for convolutive blind source separation. In:  
*Independent Component Analysis and Blind Signal Separation:  
Second International Workshop, Helsinki, Finland, Juni 2000,  
Proceedings*, 2000, S. 215–220
- [Ama97] AMARI, S.: Neural learning in structured parameter spaces –  
natural Riemannian gradient. In: *In Advances in Neural  
Information Processing Systems*, MIT Press, 1997, S. 127–133
- [AMSM05] ARAKI, S. ; MAKINO, S. ; SAWADA, H. ; MUKAI, R.: Reducing  
musical noise by a fine-shift overlap-add method applied to source  
separation using a time-frequency mask. In: *IEEE International  
Conference on Acoustics, Speech, and Signal Processing,  
Proceedings* Bd. 3, 2005, S. 81–84
- [Aos81] AOSHIMA, N.: Computer generated Pulse Signal applied for Sound  
Measurement. In: *The Journal of the Acoustical Society of America*  
69 (1981), Mai, Nr. 5, S. 1484–1488
- [AR77] ALLEN, J.B. ; RABINER, L.R.: A unified approach to short-time  
Fourier analysis and synthesis. In: *Proceedings of the IEEE* 65  
(1977), November, Nr. 1, S. 1558–1564

- [Arn04] ARNDT, Ch.: *Information Measures: Information and its Description in Science and Engineering*. 1st ed. 2001, 2nd printing. Springer Verlag, 2004 (Signals and Communication Technology). – ISBN 3-540-40855-X
- [BACM97] BELOUHRANI, A. ; ABED MERAÏM, Karim ; CARDOSO, J.-F. ; MOULINES, É.: A blind source separation technique based on second order statistics. In: *IEEE Transactions on Signal Processing* 45 (1997), Februar, Nr. 2, S. 434-444
- [Bar98] BARTHE, F.: Optimal Young's inequality and its converse: a simple proof. In: *GAFSA, Geometric And Functional Analysis* 8 (1998), August, S. 234-242
- [Bau01] BAUER, Ch.: *Independent component analysis of biomedical signals: modern data analysis techniques can assist medical decision making*, Universität Regensburg, Diss., 2001
- [Bau05] BAUMANN, W.: *Optimierung frequenzvarianter Nullbeamformer für akustische Signale mittels Statistik höherer Ordnung - Anwendungen im Kfz und in Büroräumen*, Technische Universität Berlin, Diss., 2005
- [BC94] BROWN, G. ; COOKE, M.: Computational auditory scene analysis. In: *Computer Speech and Language* 8 (1994), S. 297-336
- [BF05] BOUBCHIR, L. ; FADILI, J. M.: Multivariate statistical modeling of images with the curvelet transform. In: *IEEE Signal Processing and Its Applications, 2005. Eighth International Symposium, Proceedings* Bd. 2, 2005, S. 747-750
- [BHHK01] BEN HAMZA, A. ; HE, Yun ; KRIM, H.: An information divergence measure for ISAR image registration. In: *11th IEEE Signal Processing Workshop on Statistical Signal Processing, Proceedings*, 2001, S. 130-133
- [BKKO01] BAUMANN, W. ; KÖHLER, B.-U. ; KOLOSSA, D. ; ORGLMEISTER, R.: Real Time Separation of Convolutional Mixtures. In: *Independent Component Analysis and Blind Signal Separation: 4th International Symposium, ICA 2001, San Diego, USA, Proceedings*, 2001, S. 65-69
- [BKO03] BAUMANN, W. ; KOLOSSA, D. ; ORGLMEISTER, R.: Maximum Likelihood Permutation Correction for Convolutional Source

- Separation. In: *Independent Component Analysis and Blind Signal Separation: 4th International Symposium, ICA 2003, Nara, Japan, April 2003, Proceedings*, 2003, S. 373–378
- [BKP<sup>+</sup>08] BOLDT, J. B. ; KJEMS, U. ; PEDERSEN, M. S. ; LUNNER, T. ; WANG, D. L.: Estimation of the ideal binary mask using directional systems. In: *11th International Workshop on Acoustic Echo and Noise Control, Seattle, WA, 2008, Proceedings*, 2008
- [BR82] BURBEA, J. ; RAO, C.: On the convexity of some divergence measures based on entropy functions. In: *IEEE Transactions on Information Theory* 28 (1982), Mai, Nr. 3, S. 489–495
- [Bre90] BREGMAN, A. S.: *Auditory scene analysis*. Cambridge, MA : MIT Press, 1990
- [Bri95] BRIGHAM, E. O.: *Schnelle Fourier Transformation*. 6., korr. Auflage. Oldenbourg, 1995. – ISBN 3–486–23177–4
- [BS95a] BELL, A. J. ; SEJNOWSKI, T. J.: An Information-Maximization Approach to Blind Separation and Blind Deconvolution. In: *Neural Computation* 7 (1995), Nr. 6, S. 1129–1159
- [BS95b] BELL, A. J. ; SEJNOWSKI, T. J.: A non-linear information maximisation algorithm that performs blind separation. In: *in Advances in Neural Information Processing Systems 7*, MIT Press, 1995, S. 467–474
- [BSMM00] BRONSTEIN, I. N. ; SEMENDJAJEW, K. A. ; MUSIOL, G. ; MÜHLIG, H.: *Taschenbuch der Mathematik*. überarbeitete und erweiterte Auflage 5. Frankfurt am Main : Verlag Harri Deutsch, Thun und Frankfurt, 2000
- [BV02] BERCHER, J. F. ; VIGNAT, C.: A Rényi entropy convolution inequality with application. In: *Processings on EUSIPCO 2002, Toulouse, France, 2002*
- [CA02] CICHOCKI, A. ; AMARI, S.: *Adaptive Blind Signal and Image Processing*. New York : John Wiley, 2002. – ISBN 0–471–60791–6
- [Car03] CARDOSO, J.-F.: Dependence, Correlation and Gaussianity in Independent Component Analysis. In: *Journal of Machine Learning Research* 4 (2003), Nr. 7-8, S. 1177–1203

- [CB01] COHEN, I. ; BERDUGO, B.: Speech enhancement for non-stationary noise environments. In: *Signal Processing* 81 (2001), Oktober, Nr. 11, S. 2403–2418
- [Coh65] COHEN, A. C.: Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. In: *Technometrics* 7 (1965), November, Nr. 4, S. 579–588
- [Coh03] COHEN, I.: Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. In: *IEEE Transactions on Speech and Audio Processing* 11 (2003), Nr. 5, S. 466–475
- [Coh05] COHEN, I.: Relaxed Statistical Model for Speech Enhancement and a Priori SNR Estimation. In: *IEEE Transactions on Speech and Audio Processing* 13 (2005), Nr. 5, S. 870–881
- [CS93] CARDOSO, J.F. ; SOULOUMIAC, A.: Blind beamforming for non-Gaussian signals. In: *Radar and Signal Processing, IEEE Proceedings F* 140 (1993), Dezember, Nr. 6, S. 362–370
- [CT91] COVER, T. M. ; THOMAS, J. A.: *Elements of Information Theory*. New York : Wiley, 1991
- [DBC91] DENDRINOS, M. ; BAKAMIDIS, Stelios ; CARAYANNIS, George: Speech enhancement from noise: A regenerative approach. In: *Speech Communication* 10 (1991), Nr. 1, S. 45–57
- [DCT91] DEMBO, A. ; COVER, T.M. ; THOMAS, J.A.: Information theoretic inequalities. In: *IEEE Transactions on Information Theory* 37 (1991), November, Nr. 6, S. 1501–1518. – ISSN 0018–9448
- [DMGFRD03] DOMINGUEZ-MOLINA, J. A. ; GONZALEZ-FARIAS, G. ; RODRIGUEZ-DAGNINO, R. M.: A practical procedure to estimate the shape parameter in the generalized Gaussian distribution. (2003). [http://www.cimat.mx/reportes/enlinea/I-01-18\\_eng.pdf](http://www.cimat.mx/reportes/enlinea/I-01-18_eng.pdf)
- [Dör98] DÖRBECKER, M.: *Mehrkanalige Signalverarbeitung zur Verbesserung akustisch gestörter Sprachsignale am Beispiel elektronischer Hörhilfen*, IND, RWTH Aachen, Diss., 1998
- [DV00] DARBELLAY, G.A. ; VAJDA, I.: Entropy expressions for multivariate continuous distributions. In: *IEEE Transactions on Information Theory* 46 (2000), März, Nr. 2, S. 709–712

- [EHIP01] ERDOGMUS, D. ; HILD II, K.E. ; PRINCIPE, J.C.: Independent components analysis using Renyi's mutual information and Legendre density estimation. In: *Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on* Bd. 4. Washington, DC, USA : ACM, Juli 2001, S. 2762–2767
- [Ell06] ELLIS, D.: Model-Based Scene Analysis. In: WANG, D. (Hrsg.) ; BROWN, G. (Hrsg.): *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Wiley/IEEE Press, 2006, Kapitel 4, S. 115–146
- [EM84] EPHRAIM, Y. ; MALAH, D.: Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. In: *IEEE Transactions on Acoustics, Speech and Signal Processing* 32 (1984), Dezember, Nr. 6, S. 1109–1121
- [EM85] EPHRAIM, Y. ; MALAH, D.: Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. In: *IEEE Transactions on Acoustics, Speech and Signal Processing* 33 (1985), April, Nr. 2, S. 443–445
- [EMJ89] EPHRAIM, Y. ; MALAH, D. ; JUANG, B.-H.: Speech enhancement based upon hidden Markov modeling. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1989. ICASSP-89* Bd. 1, 1989, S. 353–356
- [EVT95] EPHRAIM, Y. ; VAN TREES, H.L.: A signal subspace approach for speech enhancement. In: *Speech and Audio Processing, IEEE Transactions on* 3 (1995), Juli, Nr. 4, S. 251–266
- [FSPZ07] FOX, B. ; SABIN, A. ; PARDO, B. ; ZOPF, A.: Modeling Perceptual Similarity of Audio Signals for Blind Source Separation Evaluation. In: *Independent Component Analysis and Signal Separation: 7th International Conference, ICA 2007, London, UK, September 2007, Proceedings* Bd. 4666. Berlin/Heidelberg : Springer-Verlag, 2007 (LNCS), S. 454–461
- [Fur06] FURUICHI, S.: Information theoretical properties of Tsallis entropies. In: *Journal of Mathematical Physics* 47 (2006), Nr. 2
- [GJV98] GUSTAFSSON, S. ; JAX, P. ; VARY, P.: A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristics. In: *Proceedings of the 1998 IEEE*

- International Conference on Acoustics, Speech and Signal Processing, 1998*. Bd. 1, 1998, S. 397–400
- [HC67] HAVRDA, J. ; CHARVAT, F.: Quantification method of classification processes: concept of structural  $\alpha$ -entropy. In: *Kybernetika* (1967), S. 30–35
- [HHK03] HE, Y. ; HAMZA, A. B. ; KRIM, H.: A generalized divergence measure for robust image registration. In: *IEEE Transactions on Signal Processing* 51 (2003), May, Nr. 5, S. 1211–1220
- [HKKO12] HOFFMANN, E. ; KOLOSSA, D. ; KÖHLER, B.-U. ; ORGLMEISTER, R.: Using information theoretic distance measures for solving the permutation problem of blind source separation of speech signals. In: *EURASIP Journal on Audio, Speech, and Music Processing* 2012:14 (2012), April
- [HKO01] HYVÄRINEN, A. ; KARHUNEN, J. ; OJA, E.: *Independent Component Analysis*. New York : John Wiley, 2001
- [HKO07a] HOFFMANN, E. ; KOLOSSA, D. ; ORGLMEISTER, R.: A Batch Algorithm for Blind Source Separation of Acoustic Signals Using ICA and Time-Frequency Masking. In: *Independent Component Analysis and Signal Separation: 7th International Conference, ICA 2007, London, UK, September 2007, Proceedings* Bd. 4666. Berlin/Heidelberg : Springer-Verlag, 2007 (LNCS), S. 480–487
- [HKO07b] HOFFMANN, E. ; KOLOSSA, D. ; ORGLMEISTER, R.: A Soft Masking Strategy based on Multichannel Speech Probability Estimation for Source Separation and Robust Speech Recognition. In: *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. New Paltz, NY, 2007
- [HKO09] HOFFMANN, E. ; KOLOSSA, D. ; ORGLMEISTER, R.: Time Frequency Masking Strategy for Blind Source Separation of Acoustic Signals Based on Optimally-Modified LOG-Spectral Amplitude Estimator. In: *Independent Component Analysis and Signal Separation: 8th International Conference, ICA 2009, Paraty, Brazil, March 2009, Proceedings* Bd. 5441. Berlin/Heidelberg : Springer Verlag, 2009 (LNCS), S. 581–588
- [HKO11] HOFFMANN, E. ; KOLOSSA, D. ; ORGLMEISTER, R. ; KOLOSSA, D. (Hrsg.) ; HAEB-UMBACH, R. (Hrsg.): *Recognition of Multiple Speech Sources using ICA*. Springer Verlag, 2011. – 319–344 S.

- [HKO12] HOFFMANN, E. ; KOLOSSA, D. ; ORGLMEISTER, R.: Time-Frequency-Processing for ICA-Supported Speech Recognition in Multitalker Conditions. In: *Proceedings DAGA, invited paper*. Darmstadt, Germany, March 2012
- [HMMG01] HERO, A. O. ; MA, B. ; MICHEL, O. ; GORMAN, J.: Alpha divergence for classification, indexing and retrieval / University of Michigan. 2001. – Forschungsbericht
- [HSK04] HASAN, M.K. ; SALAHUDDIN, S. ; KHAN, M.R.: A modified a priori SNR for speech enhancement using spectral subtraction rules. In: *Signal Processing Letters, IEEE* 11 (2004), April, Nr. 4, S. 450–453
- [HW04] HU, Guoning ; WANG, DeLiang: Monaural speech segregation based on pitch tracking and amplitude modulation. In: *IEEE Transactions on Neural Networks* 15 (2004), Nr. 5, S. 1135–1150
- [Hyv99] HYVÄRINEN, A.: Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. In: *IEEE Transactions on Neural Networks* 10 (1999), Nr. 3, S. 626–634
- [ITU03] *Recommendation ITU-R BS.1534-1: Method for the subjective assessment of intermediate quality level of coding systems*. <http://www.itu.int/rec/recommendation.asp?type=folders&parent=R-REC-BS.1534>. Version: 2003
- [JVA<sup>+</sup>08] JAFARI, M. G. ; VINCENT, E. ; ABDALLAH, S. A. ; PLUMBLEY, M. D. ; DAVIES, M. E.: An adaptive stereo basis method for convolutive blind audio source separation. In: *Neurocomputation* 71 (2008), Nr. 10-12, S. 2087–2097
- [KBM<sup>+</sup>98] KAWAMOTO, M. ; BARROS, A. K. ; MANSOUR, A. ; MATSUOKA, K. ; OHNISHI, N.: Blind separation for convolutive mixtures of non-stationary signals. In: *in Int. Conf. Neural Inf. Proc*, 1998, S. 743–746
- [KGG89] KOO, B. ; GIBSON, J.D. ; GRAY, S.D.: Filtering of colored noise for speech enhancement and coding. In: *IEEE International Conference on Acoustics, Speech, and Signal, Processing, 1989. ICASSP-89* Bd. 1, 1989, S. 349–352
- [Köh99] KÖHLER, B.-U.: *Realzeitfähige blinde Quellentrennung am Beispiel elektroenzephalographischer Signale*, Technische Universität Berlin, Diss., 1999

- [Köh05] KÖHLER, B.-U.: *Konzepte der statistischen Signalverarbeitung*. Berlin : Springer-Verlag, 2005. – ISBN 3–540–23491–8
- [KHK04] KAMATA, K. ; HU, X. ; KOBATAKE, H.: A New Approach to the Permutation Problem in Frequency Domain Blind Source Separation. In: *Independent Component Analysis and Signal Separation: Fifth International Conference, ICA 2004, Granada, Spain, September 22-24, 2004. Proceedings* Bd. 3195. Berlin/Heidelberg : Springer-Verlag, 2004 (LNCS), S. 849–856
- [KK02a] KAMMEYER, K. D. ; KROSCHER, K.: *Digitale Signalverarbeitung: Filterung und Spektralanalyse*. 5., durchges. u. erg. Auflage. Stuttgart, D : B. G. Teubner Verlag, 2002. – ISBN 3–519–46122–6
- [KK02b] KLEIN, M. ; KABAL, P.: Signal subspace speech enhancement with perceptual post-filtering. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02)*. Bd. 1, 2002, S. 537–540
- [KKEP07] KARAKOS, D. ; KHUDANPUR, S. ; EISNER, J. ; PRIEBE, C.E.: Iterative Denoising using Jensen-Renyi Divergences with an Application to Unsupervised Document Categorization. In: *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*. Bd. 2, 2007, S. II–509–II–512
- [KN04] KOKKINAKIS, K. ; NANDI, A. K.: Multichannel Speech Separation Using Adaptive Parameterization of Source PDFs. In: *Independent Component Analysis and Signal Separation: Fifth International Conference, ICA 2004, Granada, Spain, September 22-24, 2004. Proceedings* Bd. 3195. Berlin/Heidelberg : Springer-Verlag, 2004 (LNCS), S. 486–493
- [KO09] KOLOSSA, D. ; ORGLMEISTER, R.: Nonlinear Postprocessing for Blind Speech Separation. In: *Independent Component Analysis and Blind Signal Separation: Fifth International Conference, ICA 2004, Granada, Spain, September 22-24, 2004, Proceedings*. Berlin/Heidelberg : Springer-Verlag, 2009, S. 832–839
- [Kol07] KOLOSSA, D.: *Independent Component Analysis for Environmentally Robust Speech Recognition*, TU Berlin, Diss., 2007
- [KSK<sup>+</sup>00] KURITA, S. ; SARUWATARI, H. ; KAJITA, S. ; TAKEDA, K. ; ITAKURA, F.: Evaluation of blind signal separation method using

- directivity pattern under reverberant conditions. In: *Proceedings ICASSP 2000* Bd. 5, 2000, S. 3140 – 3143
- [Lam96] LAMBERT, R. H.: *Multichannel blind deconvolution: Fir matrix algebra and separation of multipath mixtures*, University of Southern California, Department of Electrical Engineering, Diss., 1996
- [LBL97] LEE, T.-W. ; BELL, A. J. ; LAMBERT, R. H.: Blind separation of delayed and convolved sources. In: *Advances in Neural Information Processing Systems* Bd. 9. Cambridge, MA : MIT Press, 1997, S. 758–764
- [LDC93] LDC: *TIDigits Speech Database: Studio Quality Speaker-Independent Connected-Digit Corpus*. <http://morph ldc.upenn.edu/Catalog/LDC93S10.html>. Version: 1993
- [LL00] LEE, T.-W. ; LEWICKI, M. S.: The Generalized Gaussian Mixture Model Using ICA. In: *Independent Component Analysis and Blind Signal Separation: Second International Workshop, Helsinki, Finland, Juni 2000, Proceedings*, 2000, S. 239–244
- [LO79] LIM, J.S. ; OPPENHEIM, A.V.: Enhancement and bandwidth compression of noisy speech. In: *Proceedings of the IEEE* 67 (1979), Dezember, Nr. 12, S. 1586–1604
- [Loi07] LOIZOU, P. C.: *Speech Enhancement: Theory and Practice (Signal Processing and Communications)*. 1. CRC, 2007. – ISBN 0849350328
- [LPP07] LIU, W. ; POKHAREL, P.P. ; PRINCIPE, J.C.: Correntropy: Properties and Applications in Non-Gaussian Signal Processing. In: *IEEE Transactions on Signal Processing* 55 (2007), November, Nr. 11, S. 5286–5298
- [Lyo83] LYON, R.: A computational model of binaural localization and separation. In: *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '83*. Bd. 8, 1983, S. 1148–1151
- [MD03] MITIANOUDIS, N. ; DAVIES, M. E.: Audio Source Separation of Convolutional Mixtures. In: *IEEE Transactions on Audio and Speech Processing* 11 (2003), Nr. 5, S. 489–497
- [MDC00] MEJUTO, C. ; DAPENA, A. ; CASTEDO, L.: Frequency-domain infomax for blind separation of convolutional mixtures. In:

- Independent Component Analysis and Blind Signal Separation: Second International Conference, ICA 2000, Proceedings, 2000*, S. 315–320
- [Mer09] MERCER, J.: Functions of positive and negative type and their connection with the theory of integral equations. In: *Philosophical Transactions of the Royal Society, London A* (1909), Nr. 209, S. 415–446
- [MFA<sup>+</sup>08] MARTINS, A. F. T. ; FIGUEIREDO, M. A. T. ; AGUIAR, P. M. Q. ; SMITH, N. A. ; XING, E. P.: Nonextensive entropic kernels. In: *ICML '08: 25th international conference on Machine learning, Proceedings* Bd. 307. Helsinki, Finland : ACM, 2008, S. 640–647
- [MIZ01] MURATA, N. ; IKEDA, S. ; ZIEHE, A.: An approach to blind source separation based on temporal structure of speech signals. In: *Neurocomputing* 41 (2001), S. 1–24
- [MK03] MANSOUR, A. ; KAWAMOTO, M.: ICA Papers Classified According to their Applications and Performances. In: *IEICA Transactions on Fundamentals* E86-A. No. 3 (2003), März, S. 620–633
- [MM80] MCAULAY, R. ; MALPASS, M.: Speech enhancement using a soft-decision noise suppression filter. In: *IEEE Transactions on Acoustics, Speech and Signal Processing* 28 (1980), April, Nr. 2, S. 137–145
- [MM07] MAZUR, Radoslaw ; MERTINS, Alfred: Solving the Permutation Problem in Convolutional Blind Source Separation. In: *Independent Component Analysis and Signal Separation: 7th International Conference, ICA 2007, London, UK, September 2007, Proceedings* Bd. 4666. Berlin/Heidelberg : Springer-Verlag, 2007 (LNCS), S. 512–519
- [MSAM03] MUKAI, R. ; SAWADA, H. ; ARAKI, S. ; MAKINO, S.: Real-Time Blind Source Separation for Moving Speakers using Blockwise ICA and Residual Crosstalk Subtraction. In: *Independent Component Analysis and Blind Signal Separation: 4th International Symposium, ICA 2003, Nara, Japan, April 2003, Proceedings, 2003*, S. 975–980
- [NP83] NIKIAS, C. L. ; PETROPULU, A. P.: *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*. New Jersey : Prentice Hall, 1983

- [NP94] NADAL, J.-P. ; PARGA, N.: Non linear neurons in the low noise limit: a factorial code maximizes information transfer. In: *Network* 5 (1994), Nr. 4, S. 565–581
- [Pet02] PETERS, M.: *Psychoakustische Signalverbesserung und Geräuschreduktion in Kraftfahrzeugen*, Universität Kaiserslautern, Elektrische Informationstechnik, Diss., 2002
- [PLKP07] PEDERSEN, M. S. ; LARSEN, J. ; KJEMS, U. ; PARRA, L. C.: A Survey of Convolutional Blind Source Separation Methods. In: *Springer Handbook of Speech Processing*. Springer, November 2007
- [Pra05] PRASAD, R.: *Fixed-Point ICA based Speech Signal Separation and Enhancement with Generalized Gaussian Model*, Department of Information Science Graduate School of Information Science Nara Institute of Science and Technology, Diss., 2005. [citeseer.ist.psu.edu/prasad05fixedpoint.html](http://citeseer.ist.psu.edu/prasad05fixedpoint.html)
- [PS00] PARRA, L. ; SPENCE, C.: Convolutional Blind Source Separation of Non-Stationary Sources. In: *IEEE Transactions on Speech and Audio Processing*, 2000, S. 320–327
- [PSB03a] PHAM, D.-T. ; SERVIÈRE, C. ; BOUMARAF, H.: Blind Separation of Convolutional Audio Mixtures using Nonstationarity. In: *Independent Component Analysis and Blind Signal Separation: 4th International Symposium, ICA 2003, Nara, Japan, April 2003, Proceedings*, 2003, S. 981–986
- [PSB03b] PHAM, D.-T. ; SERVIÈRE, C. ; BOUMARAF, H.: Blind Separation of Speech Mixtures Based on Nonstationarity. In: *7th International Symposium on Signal Processing and Its Applications, ISSPA '03, Paris, France, Proceedings Bd. 2*, 2003, S. 73–76
- [PXF00] PRINCIPE, J. C. ; XU, D. ; FISHER, J. W.: Informatin-theoretic learning. In: *Unsupervised Adaptive Filtering* (2000), S. 265–320
- [R76] RÉNYI, A.: On measures of entropy and information. In: *Selected Papers of Alfred Rényi* Bd. 2. Budapest : Akademia Kiado, 1976, S. 565–580
- [RABH07] ROBLEDO-ARNUNCIO, E. ; BIING-HWANG, J.: Blind Source Separation of Acoustic Mixtures with Distributed Microphones. In: *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007*. 3 (2007), April, S. 949–952

- [RAJ07] ROBLEDO-ARNUNCIO, E. ; JUANG, Biing-Hwang: Blind Source Separation of Acoustic Mixtures with Distributed Microphones. In: *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*. Bd. 3, 2007, S. III-949-III-952
- [Ric45] RICE, S. O.: Mathematical Analysis of Random Noise. In: *Bell System Technical Journal* 24 (1945), S. 46-156
- [Row00] ROWEIS, S. T.: One Microphone Source Separation. In: *Advances in Neural Information Processing Systems*. Cambridge, MA : MIT Press, 2000, S. 793-799
- [RR07] REDDY, A.M. ; RAJ, B.: Soft Mask Methods for Single-Channel Speaker Separation. In: *IEEE Transactions on Audio, Speech, and Language Processing* 15 (2007), August, Nr. 6, S. 1766-1776
- [RW04] ROMAN, N. ; WANG, DeLiang: Binaural sound segregation for multisource reverberant environments. In: *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on* Bd. 2, 2004, S. ii-373-6
- [SAM07] SAWADA, H. ; ARAKI, S. ; MAKINO, S.: Measuring Dependence of Bin-wise Separated Signals for Permutation Alignment in Frequency-domain BSS. In: *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on* (2007), Mai, S. 3247-3250
- [SAMM06] SAWADA, H. ; ARAKI, S. ; MUKAI, R. ; MAKINO, S.: Blind Extraction of Dominant Target Sources Using ICA and Time-Frequency Masking. In: *IEEE Transactions on Audio, Speech and Language Processing, [see also IEEE Transactions on Speech and Audio Processing]* 14 (2006), Nr. 6, S. 2165-2173. – ISSN 1558-7916
- [Saw03] SAWADA, H.: *BSS Sound Demonstration*. <http://www.kecl.ntt.co.jp/icl/signal/sawada/demo/bss2to4/index.html>.  
Version: 2003
- [Sch97] SCHWARZ, H. R.: *Numerische Mathematik*. Stuttgart : B. G. Tuebner, 1997. – ISBN 3-519-32960-3
- [Sch01] SCHOBEN, D. W.: *Real-Time Adaptive Concepts in Acoustics*. Dordrecht : Kluwer Academic Publishers, 2001
- [SDSVD98] SIJBERS, J. ; DEKKER, A.J. den ; SCHEUNDERS, P. ; VAN DYCK, D.: Maximum-likelihood estimation of Rician distribution

- parameters. In: *IEEE Transactions on Medical Imaging* 17 (1998), Juni, Nr. 3, S. 357–361
- [SF96] SCALART, P. ; FILHO, J. V.: Speech enhancement based on a priori signal to noise estimation. In: *ICASSP '96: Acoustics, Speech, and Signal Processing, 1996. on Conference Proceedings., 1996 IEEE International Conference, Proceedings.* Washington, DC, USA : IEEE Computer Society, 1996. – ISBN 0-7803-3192-3, S. 629–632
- [SG07] SUN, S. ; GUO, C.: Image Registration by Minimizing Tsallis Divergence Measure. In: *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on* Bd. 4, 2007, S. 712–715
- [SG09] SUDHAKAR, P. ; GRIBONVAL, R.: A Sparsity-Based Method to Solve Permutation Indeterminacy in Frequency-Domain Convolutional Blind Source Separation. In: *Independent Component Analysis and Signal Separation: 8th International Conference, ICA 2009, Paraty, Brazil, March 2009, Proceedings* Bd. 5441. Berlin/Heidelberg : Springer-Verlag, 2009 (LNCS), S. 338–345
- [Sil86] SILVERMAN, B. W.: *Density Estimation for Statistics and Data Analysis.* Chapman & Hall/CRC, 1986. – ISBN 0412246201
- [SK00] SOON, I.Y. ; KOH, S.N.: Low distortion speech enhancement. In: *Vision, Image and Signal Processing, IEEE Proceedings on* 147 (2000), Juni, Nr. 3, S. 247–253
- [Sma97] SMARAGDIS, P.: Efficient blind separation of convolved sound mixtures. In: *Applications of Signal Processing to Audio and Acoustics, 1997. 1997 IEEE ASSP Workshop on,* 1997
- [SMAM04] SAWADA, H. ; MUKAI, R. ; ARAKI, S. ; MAKINO, S.: A robust and precise method for solving the permutation problem of frequency-domain blind source separation. In: *IEEE Transactions on Speech and Audio Processing* 12 (2004), Nr. 5, S. 530–538
- [SMM03] SAWADA, H. ; MUKAI, R. ; MAKINO, S.: Direction of Arrival estimation for Multiple Source Signals using Independent Component Analysis. In: *7th International Symposium on Signal Processing and Its Applications, ISSPA '03, Proceedings* Bd. 2, 2003, S. 411–414

- [SPP06] SANTAMARIA, I. ; POKHAREL, P.P. ; PRINCIPE, J.C.: Generalized correlation function: definition, properties, and application to blind equalization. In: *IEEE Transactions on Signal Processing* 54 (2006), Juni, Nr. 6, S. 2187–2197
- [SS01] SCHÖLKOPF, B. ; SMOLA, A.J.: *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA : MIT Press, 2001. – ISBN 0262194759
- [STS99] SCHOBEN, Daniel ; TORKKOLA, Kari ; SMARAGDIS, Paris: Evaluation Of Blind Signal Separation Methods. In: *Independent Component Analysis and Signal Separation: First International Workshop, ICA 1999, Aissios, France, January 1999, Proceedings*, 1999, S. 261–266
- [Tre68] TREES, H. L. V.: *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. New York : John Wiley, 1968
- [Tre02] TREES, H. L. V.: *Detection, Estimation, and Modulation Theory, Part IV: Optimum Array Processing*. John Wiley, 2002. – ISBN 0-471-09390-4
- [Tsa88] TSALLIS, Constantino: Possible generalization of Boltzmann-Gibbs statistics. In: *Journal of Statistical Physics* 52 (1988), Juli, Nr. 1-2, S. 479–487
- [Tuk] TUKEY, J. W.: *Exploratory data analysis*. Addison-Wesley. – ISBN 0-201-07616-0
- [VA89] VARANASI, M. K. ; AAZHANG, B.: Parametric generalized Gaussian density estimation. In: *Journal of the Acoustical Society of America* 86 (1989), Oktober, Nr. 4, S. 1404–1415
- [VHH98] VARY, P. ; HEUTE, U. ; HESS, W.: *Digitale Sprachsignalverarbeitung*. Stuttgart : B.G. Teubner, 1998
- [Vin05] VINCENT, E.: *MUSHRAM 1.0: User guide*. Centre for Digital Music Queen Mary, University of London, November 2005. <http://www.elec.qmul.ac.uk/people/emmanuelv/mushram/mushram.zip>
- [VJP06] VINCENT, E. ; JAFARI, M. G. ; PLUMBLEY, M. D.: Preliminary guidelines for subjective evaluation of audio source separation algorithms. In: NANDI, A. K. (Hrsg.) ; ZHU, X. (Hrsg.):

- 
- Proceedings of the ICA Research Network International Workshop, Liverpool, UK, 18-19 Sept 2006*, 2006, S. 93–96
- [VL03] VISSER, E. ; LEE, T.-W.: Application of Blind Source Separation in Speech Processing for Combined Interference Removal and Robust Speaker Detection using a Two-Microphone Setup. In: *Independent Component Analysis and Blind Signal Separation: 4th International Symposium, ICA 2003, Nara, Japan, April 2003, Proceedings*, 2003, S. 325–329
- [WB06] WANG, D. (Hrsg.) ; BROWN, Guy J. (Hrsg.): *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Hoboken, NJ : Wiley-IEEE Press, 2006. – ISBN 0471741094
- [WL82] WANG, D. ; LIM, Jae: The unimportance of phase in speech enhancement. In: *IEEE Transactions on Acoustics, Speech and Signal Processing* 30 (1982), August, Nr. 4, S. 679–681
- [Yam02] YAMANO, Takuya: Some properties of q-logarithm and q-exponential functions in Tsallis statistics. In: *Physica A: Statistical Mechanics and its Applications* 305 (2002), März, Nr. 3-4, S. 486–496
- [YFD99] YASMIN, A. ; FIEGUTH, P. ; DENG, Li: Speech enhancement using voice source models. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 1999, Proceedings*. Washington, DC, USA : IEEE Computer Society, 1999, S. 797–800
- [YR04] YILMAZ, Ö. ; RICKARD, S.: Blind Separation of Speech Mixtures via Time-Frequency Masking. In: *IEEE Transactions on Signal Processing* 52 (2004), Juli, Nr. 7, S. 1830–1847
- [ZN05] ZOGRAFOS, K. ; NADARAJAH, S.: Expressions for Rényi and Shannon entropies for multivariate distributions. In: *Statistics & Probability Letters* 71 (2005), Januar, Nr. 1, 71-84. <http://ideas.repec.org/a/eee/stapro/v71y2005i1p71-84.html>